UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

New insights on black rot of crucifers: disclosing novel virulence genes by in vivo host/pathogen transcriptomics and functional genetics

Doutoramento em Biologia
Microbiologia

Joana Costa Cardoso da Cruz

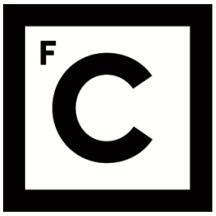
Tese orientada por:
Doutora Maria Leonor Pato da Cruz
Professor Doutor Rogério Paulo de Andrade Tenreiro

Documento especialmente elaborado para a obtenção do grau de doutor

2017

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS



Ciências
ULisboa

New insights on black rot of crucifers: disclosing novel virulence genes by in vivo host/pathogen transcriptomics and functional genetics

Doutoramento em Biologia

Microbiologia

Joana Costa Cardoso da Cruz

Tese orientada por:

Doutora Maria Leonor Pato da Cruz

Professor Doutor Rogério Paulo de Andrade Tenreiro

Júri:

Presidente:

- Doutor Rui Manuel Dos Santos Malhó, Professor Catedrático, Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor Pedro Miguel da Silva Albuquerque, Bolseiro Pós-Doutoramento, Instituto de Biologia Molecular e Celular
- Doutora Maria Teresa Ferreira de Oliveira Barreto Goulão Crespo, Responsável de Laboratório – Investigadora Nível 3, Instituto de Biologia Experimental e Tecnológica (IBET);
- Doutora Amélia Maria Pereira Lopes, Investigadora Auxiliar, Instituto Nacional de Investigação Agrária e Veterinária (INIAV);
- Doutora Maria Leonor Pato da Cruz, Investigadora Auxiliar, Instituto Nacional de Investigação Agrária e Veterinária (INIAV) (Orientadora);
- Doutora Maria Helena Mendes da Costa Ferreira Correia de Oliveira, Professora Associada, Instituto Superior de Agronomia da Universidade de Lisboa
- Doutor Manuel Pedro Salema Fevereiro, Professor Auxiliar com Agregação, Faculdade de Ciências da Universidade de Lisboa;
- Doutora Margarida Henriques da Gama Carvalho, Professora Auxiliar, Faculdade de Ciências da Universidade de Lisboa;

Documento especialmente elaborado para a obtenção do grau de doutor

Bolsa Individual de Doutoramento SFRH/BD/69787/2010
Fundação para a Ciência e a Tecnologia

2017

Acknowledgments

This thesis is the result of a life-changing experience, an adventure down a sinuous path that I would most certainly could not have walked alone. During this journey, I was always welcomed, encouraged and supported by all the people around me and for that I will be forever thankful.

First and foremost, I would like to thank my supervisor, Dr. Leonor Cruz, head of the Phytobacteriology Lab (INIAV), without whom this work would not have been achievable. I am thankful for your scientific guidance, support and mentoring. For allways inspiring me, believing in me and encouraging me to challenge myself. But above all, I am deeply grateful for your selfless dedication and true friendship, that led me every step of the way.

I would also like to extend my profound gratitude to my co-supervisor, Professor Rogério Tenreiro, head of the Microbiology & Biotechnology Group (BioISI/FCUL), for his constant scientific leadership, constructive criticism and advice, pushing me to the limit and beyond, relentlessly motivating me to do more and better. For that, you will always be *The Professor*.

To both my supervisors, I would like to say that you have made me a better scientist and a better person. I hope our paths continue to cross.

I would like to show my appreciation to the head of UEIS-SAFSV, Dr. Amélia Lopes, for kindly welcoming me at INIAV ever since I was a MSc student, making this institution my home away from home.

To the team of the GMO Lab (INIAV), thank you for your support during the experimental work related to the last chapter of this thesis.

To Dr. Raquel Sá-Leão and Dr. Carina Valente, at the Molecular Microbiology of Human Pathogens Lab (ITQB), for their kind collaboration during the use of their Bioanalyzer System.

To Dr. Margarida Gama-Carvalho and Dr. Andreia Fonseca, at the Gene Expression & Regulation Group (BioISI/FCUL), I would like to thank you for welcoming me and for your guidance and support during the bioinformatic analysis of the results presented in the last chapter of this thesis. You have introduced me to a new world!

To Professor Ana Tenreiro, at the Microbiology & Biotechnology Group (BioISI/FCUL), thank you for your kind collaboration and advice throughout this work.

To my colleagues at the Phytobacteriology Lab (INIAV), Isaura Velez and Camila Fernandes, and to Márcia Santos, Ana Chegão and other colleagues: thank you for your friendship and kind words. And for all the cakes and chocolates, they really kept me going.

To my childhood and university friends, thank you for your long-lasting and steady presence in my life, for sharing with me those life-giving laughs and for putting up with me during venting moments.

Last, but definitely not least, I would like to dedicate this thesis to my family. Thank you for everything, you are my reason for living.

Abstract

Xanthomonas campestris belongs to the gamma subdivision of Proteobacteria and is the type species of a genus comprising 28 species of plant pathogenic bacteria, affecting 124 species of monocotyledonous and 268 species of dicotyledonous plants.

Identified for the first time in 1895 in the United States of America, this species has undergone several taxonomic reclassifications, and currently comprises pathovars *X. campestris* pv. *campestris* (*Xcc*), *X. campestris* pv. *raphani* (Xcr) and *X. campestris* pv. *incanae* (*Xci*), causing distinct diseases in vegetables, ornamentals and spontaneous plants belonging to *Brassica*ceae family, as well as some vegetable crops belonging to the Solanaceae family.

Black rot disease, caused by *Xcc*, is the most important bacterial disease of *Brassica*ceae, affecting crops and weeds worldwide. After penetration through hydathodes on leaf margins, bacterial multiplication causes typical V-shaped lesions on leaf margins, followed by darkening of the veins that evolve to necrosis of the affected tissue, ultimately leading to plant death. Short-distance dispersal through plant-to-plant contact, human manipulation, wind, insects, aerosols and irrigation water or rain, associated with long distance dispersal through infected seeds and plantlets following commercial routes across the globe are responsible for the worldwide distribution of this disease.

Studies on the interaction of *Xcc* with *B. oleracea* have identified several virulence genes, however no resistance genes have been successfully cloned in *Brassica*ceae crops and there is still a lack of effective black rot disease control measures, which continues to cause severe economic losses worldwide. Despite the growing body of work on this subject, molecular mechanisms of host-pathogen interaction have been mostly inferred using *in vitro* approaches, resulting in a knowledge gap concerning the *in vivo* behavior of this pathogen during its interaction with plant hosts.

In Portugal, an important centre of *Brassica*ceae domestication, *Xcc* has long been identified and Portuguese *Xcc* strains have already been described as a unique sub-population of this pathogen. In this context, with the goal of bringing new insights on the molecular mechanisms of host/pathogen interaction, a set of 33 *X. campestris* strains collected in the country was characterized, in terms of pathogenicity, virulence, population structure and phylogenetic diversity. Furthermore, the *in planta* transcriptomes of the *Xcc* strains representing the extremes of the virulence spectrum during the infection process on two selected hosts were profiled, in a total of four pathosystems.

A high level of phenotypic diversity, supported by phylogenetic data, was found among *X. campestris* isolates, allowing the identification of *Xcc* closely related pathovars *Xcr* and *Xci* for the first time in Portugal. Moreover, among *Xcc* isolates, presence of races 4, 6 and 7 was recorded, and

two novel races of this pathovar, race 10 and race 11, were also described. Contrastingly, the partial virulence profiles determined by the presence of known virulence genes were highly conserved among the set of *X. campestris* strains. The integration of Portuguese strains in a global dataset comprising a total of 75 *X. campestris* strains provided a snapshot of the worldwide *X. campestris* phylogenetic diversity and population structure, correlating the existing pathovars with three distinct genetic lineages.

The identification of an intermediate link between *Xcc* and *Xcr*, suggests that these pathovars are more closely related to each other than to *Xci*. This gradient of genetic relatedness seems to be associated to the host range of each pathovar. While *Xci* appears only to be pathogenic on ornamental *Brassica*ceae, *Xcc* and *Xcr* have a partially overlapping host range: *Xcc* affects mostly *Brassica*ceous hosts, whereas *Xcr*, while affecting the same hosts, is also pathogenic on Solanaceous hosts. These findings suggest that instead of causing a host shift, the genetic divergence between *Xcc* and *Xcr* conferred the latter strains the ability to explore additional hosts, resulting in a broader host range. Although population displayed a major clonal structure, the presence of recombinational events that may have driven the ecological specialization of *X. campestris* and distinct host ranges was highlighted. Portuguese *X. campestris* strains provided a significant input of genetic diversity, confirming this region as an important diversification reservoir, most likely taking place through host-pathogen co-evolution.

Virulence assessment of Portuguese *Xcc* strains, based on the percentage on infected leaf area after inoculation, highlighted that virulence was not homogeneous within races and that higher pathogenicity was not necessarily correlated higher virulence. It was then possible to select the extremes of the virulence spectrum – CPBF213, the lowest virulence strain (L-vir), and CPBF278, the highest virulence strain (H-vir).

The *in vivo* transcriptome profiling of those contrastingly virulent strains infecting two cultivars of *B. oleracea*, using RNA-Seq, allowed establishing that *Xcc* undergoes transcriptional reprogramming in a host-independent manner, suggesting that virulence is an intrinsic feature of the pathogen. A total of 154 differentially expressed genes (DEGs) were identified between the two strains. The most represented functional category of DEGs was, as expected, 'pathogenicity and adaptation', representing 16% of the DEGs, although the complex adaptation of pathogen cells to the host environment was highlighted by the presence of DEGs from various functional categories. Among DEGs, Type III effector coding genes *xopE2* and *xopD* were induced in L-vir strain, while *xopAC*, *xopX* and *xopR* were induced in H-vir strain. In addition to Type III effectors, genes encoding proteins involved in signal transduction systems, transport, detoxification mechanisms and other virulence-associated processes were found differentially expressed between both strains, highlighting their role in virulence regulation.

Overall, low virulence appears to be the combined result of impaired sensory mechanisms, reduced detoxification of reactive oxygen species, decreased motility and higher production of pathogen-associated molecular patterns (PAMPs), accompanied by an overexpression of avirulence proteins and a repression of virulence proteins targeting the hosts' PAMP-triggered immune responses. Contrastingly, the highly virulent strain showed to be better equipped to escape initial plant defenses, whether avoiding detection, or by the ability to counteract those responses. On the other hand, upon detection, highly virulent pathogens will show a decreased expression of avirulence proteins, making them less recognizable by the hosts Effector-Triggered Immunity mechanisms, and thus able to continue multiplying and cause more severe disease symptoms.

Through the study of differential infections in planta, the highly innovative strategy used in this work contributed to disclosure of novel virulence related genes in *X. campestris* pv. *campestris* - *B. oleracea* pathosystem, that will be crucial to further detail the virulence regulation network, and develop new tools for the control of black rot disease.

Key-words

Xanthomonas campestris; population structure; pathogenicity and virulence; transcriptome profiling; RNA-Seq

Resumo

Xanthomonas campestris pertence à subdivisão gama de Proteobacteria, e é considerada a espécie tipo de um género muito diverso que compreende 28 espécies de bactérias fitopatogénicas, afectando 124 espécies de plantas monocotiledóneas e 268 espécies de plantas dicotiledóneas. Identificada pela primeira vez em 1895 nos Estados Unidos da América, esta espécie sofreu várias reclassificações taxonómicas, sendo que, actualmente, inclui as patovares *X. campestris* pv. *campestris*, *X. campestris* pv. *raphani* e *X. campestris* pv. *incanae*, causando doenças distintas em plantas hortícolas, ornamentais e espontâneas da família *Brassica*ceae, bem como em algumas culturas hortícolas da família Solanaceae.

A bacteriose vascular das crucíferas, causada por *X. campestris* pv. *campestris*, é a doença bacteriana mais importante de *Brassica*ceae, afetando especialmente diversas variedades de *Brassica oleracea*. Após penetração inicial através dos hidátodos das folhas, a multiplicação bacteriana nos vasos xilémicos provoca lesões típicas em forma de V nas margens das mesmas, acompanhadas pelo escurecimento das nervuras, culminando na necrose do tecido afetado, e provocando, em última instância, a morte da planta. A dispersão do agente patogénico a curtas distâncias ocorre maioritariamente através do contacto entre plantas, pela acção de agentes ambientais como o vento, aerossóis e insectos, bem como associada a práticas culturais desadequadas. Por outro lado, a dispersão a longas distâncias deve-se à disseminação de sementes e plântulas infectadas através de rotas comerciais. Globalmente, estes factores contribuem para que a bacteriose vascular das crucíferas constitua um problema à escala global.

Diversos estudos sobre a interação de *X. campestris* pv. *campestris* com *B. oleracea*, e com a planta modelo *Arabidopsis thaliana*, permitiram a identificação de vários genes responsáveis pela patogenicidade, bem como pela sua regulação. No entanto, nenhum gene de resistência foi clonado com sucesso em culturas de *Brassica*ceae, existindo ainda uma falta de medidas efetivas de controlo da doença, que continua a causar graves perdas económicas em todo o mundo. Apesar do número crescente de trabalhos sobre este assunto, os mecanismos moleculares que conduzem a interação patógeneo-hospedeiro têm sido maioritariamente inferidos usando abordagens *in vitro*, o que resulta numa lacuna de conhecimento no que diz respeito ao comportamento *in vivo* deste patógeneo, durante sua interação com o hospedeiro.

Em Portugal, um importante centro de domesticação de *Brassica*ceae, a presença de *X. campestris* foi há muito registada e as estirpes portuguesas desta espécie foram já descritas como uma subpopulação singular deste patógeneo. Neste contexto, com o objetivo de trazer novos conhecimentos sobre os mecanismos moleculares de interação patógeneo-hospedeiro, um conjunto de 33 estirpes de *X. campestris* isoladas no país foi caracterizada, em termos de patogenicidade,

virulência, estrutura populacional e diversidade filogenética. Adicionalmente, os transcriptomas *in planta* de duas estirpes diferencialmente virulentas de *X. campestris* pv. *campestris* foram determinados durante o processo de infecção em dois hospedeiros, num total de quatro patossistemas.

Um elevado nível de diversidade fenotípica, apoiado por dados filogenéticos, foi encontrado entre os isolados *X. campestris*, permitindo a identificação das patovares *X. campestris* pv. *raphani* e *X. campestris* pv. *incanae* pela primeira vez em Portugal. Entre os isolados de *X. campestris* pv. *campestris*, foi registada a presença das raças 4, 6 e 7, bem como de duas novas raças deste patovar, raças 10 e 11, foram também descritas. Por outro lado, os perfis parciais de virulência, determinados pela presença de genes de virulência conhecidos, foram altamente conservados entre o conjunto de estirpes de *X. campestris*, sendo que, para a maioria das estirpes, todos os genes de virulência estavam presentes. A integração das estirpes portuguesas num conjunto de dados globais, que inclui um total de 75 estirpes de *X. campestris* permitiu obter uma perspectiva da diversidade filogenética e da estrutura populacional de *X. campestris*, correlacionando os patovares existentes com três linhagens genéticas distintas.

A identificação de um elo intermediário entre *X. campestris* pv. *campestris* e *X. campestris* pv. *raphani* sugere que estes patovares estão mais próximos entre si do que de *X. campestris* pv. *incanae*. Esta proximidade genética parece estar associada à sobreposição parcial do leque de hospedeiros de *X. campestris* pv. *campestris* e X. *campestris* pv. *raphani*: *X. campestris* pv. *campestris* apenas afecta hospedeiros de *Brassica*ceae, enquanto *X. campestris* pv. *raphani* afecta também plantas da família Solanaceae. Assim, em vez de provocar uma alteração do leque de hospedeiros, a divergência genética entre estes dois patovares parece ter conferido a *X. campestris* pv. *raphani* a capacidade de explorar hospedeiros alternativos, conferindo-lhe uma gama de hospedeiros mais diversificada. Embora a população global de *X. campestris* exiba uma estrutura maioritariamente clonal, a presença de eventos de recombinação pode ter orientado a especialização ecológica de *X. campestris* com diferentes leques de hospedeiros. Numa perspectiva geral, as estirpes portuguesas de *X. campestris* facultaram um 'input' de diversidade genética significativo, confirmando esta região como um importante reservatório de diversificação, provavelmente através de um processo de co-evolução entre patógeneo e hospedeiro.

A determinação da virulência das estirpes portuguesas de *X. campestris* pv. *campestris*, com base na percentagem de área foliar afetada após a inoculação, revelou variabilidade na virulência de estirpes da mesma raça, e que uma maior patogenicidade não está necessariamente correlacionada com maior virulência. Foi então possível selecionar as estirpes representando os extremos do espectro de virulência - CPBF213, a estirpe de menor virulência (L-vir) e CPBF278, a estirpe de maior virulência (H-vir).

O transcriptoma *in vivo* das estirpes diferencialmente virulentas, determinado por RNA-Seq em duas cultivares de *B. oleracea*, permitiu estabelecer que a reprogramação transcricional de *X. campestris* pv. *campestris* durante o processo de infecção ocorre de forma não dependente do hospedeiro, sugerindo que a virulência é uma característica intrínseca do patógeneo. Foram identificados um total de 154 genes diferencialmente expressos (GDEs) entre as duas estirpes. A categoria funcional mais representada foi, como esperado, "patogenicidade e adaptação" representando 16% dos GDEs identificados, embora a complexidade da adaptação das células bacterianas ao hospedeiro tenha sido destacada pela presença de GDEs de várias categorias funcionais. Entre os GDEs, os genes codificantes para os efectores de Tipo III XopE2 e XopD estavam induzidos na estirpe de menor virulência, enquanto os que codificam para os efectores XopAC, XopX e XopR estavam induzidos na estirpe de maior virulência. Além dos efectores de Tipo III, os genes que codificam proteínas envolvidas em sistemas de transdução de sinal, transporte, mecanismos de destoxificação e outros processos associados à virulência foram encontrados diferencialmente expressos entre ambas as estirpes, evidenciando o seu papel na regulação da virulência.

Na generalidade, a baixa virulência parece ser o resultado combinado de mecanismos sensoriais enfraquecidos, reduzida destoxificação de espécies reativas de oxigénio, motilidade diminuída e maior produção de padrões moleculares associados a patógeneos, acompanhada por uma sobre-expressão de proteínas de avirulência e uma repressão de proteínas de virulência que têm como alvo as respostas imunes desencadeadas pelo hospedeiro. Por outro lado, a estirpe altamente virulenta parece estar melhor equipada para escapar às defesas iniciais da planta, evitando a detecção ou neutralizando essas respostas. Após a detecção inicial, os agentes patogénicos altamente virulentos mostram uma diminuição da expressão de proteínas de avirulência, tornando-os menos reconhecíveis pelos mecanismos de imunidade desencadeada por efectores. Desta forma, as estirpes altamente virulentas são capazes de continuar a multiplicar-se, causando uma maior severidade de sintomas.

Através do estudo de infecções diferenciais *in planta*, a estratégia altamente inovadora usada neste trabalho permitiu a identificação de novos genes relacionados com a virulência no patossistema *X. campestris* pv. *campestris* - *B. oleracea*, cruciais para detalhar os mecanismos de regulação da virulência e desenvolver novas ferramentas para o controlo da bacteriose vascular das crucíferas.

Palavras-chave

Xanthomonas campestris; estrutura populacional; patogenicidade e virulência; transcriptoma; RNA-Seq.

Contents

List of Figures

I. Introduction

II. Characterization of *Xanthomonas campestris* isolates collected in Portugal

III. *In vivo* transcriptome profiling of *Xanthomonas campestris* pv. *campestris*

List of Tables

I. Introduction

II. Characterization of *Xanthomonas campestris* isolates collected in Portugal

III. *In vivo* transcriptome profiling of *Xanthomonas campestris* pv. *campestris*

List of Abbreviations

AFLP – **A**mplified **F**ragment **L**ength **P**olymorphism

ATCC – **A**merican **T**ype **C**ulture **C**ollection

BLAST – **B**asic **L**ocal **A**lignment **S**earch **T**ool

CDS – **C**oding **D**NA **S**equence

CFBP – **C**ollection **F**rançaise de **B**actéries associées aux **P**lantes

CPBF – **C**oleção **P**ortuguesa de **B**actérias **F**itopatogénicas

DAI – **D**ays **A**fter **I**noculation

DAMP – **D**anger **A**ssociated **M**olecular **P**attern

DEG – **D**ifferentially **E**xpressed **G**enes

DSF – **D**iffusible **S**ignaling **F**actor

ETI – **E**ffector-**T**riggered **I**mmunity

ETS – **E**ffector-**T**riggered **S**usceptibility

HK – **H**istidine **K**inase

HR – **H**ypersensitive **R**esponse

IS – **I**nsertion **S**equence

ISTA – **I**nternational **S**eed **T**esting **A**ssociation

LPS – **L**ipo**p**oly**s**accharide

LRR – **L**eucin **R**ich **R**epeat

MAMP – **M**icrobe **A**ssociated **M**olecular **P**attern

MAPK - **M**itogen-**A**ctivated **P**rotein **K**inases

MLSA – **M**ulti **L**ocus **S**equence **A**nalysis

MLST – **M**ulti **L**ocus **S**equence **T**yping

NCBI – **N**ational **C**enter for **B**iotechnology **I**nformation

NCPPB – **N**ational **C**ollection of **P**lant **P**athogenic **B**acteria

NGS – **N**ext **G**eneration **S**equencing

PAMP – **P**athogen **A**ssociated **M**olecular **P**attern

PGN – **P**eptido**g**lyca**n**

PRR – **P**attern **R**ecognition **R**eceptors

PTI – **P**AMP-**T**riggered **I**mmunity

PVP – **P**artial **v**irulence **p**rofile

QS – **Q**uorum **S**ensing

QTL – **Q**uantitative **T**rait **L**oci

Rep-PCR – **Rep**etitive element sequence-based **PCR**

RH – **R**elative **H**umidity

RK – **R**eceptor **K**inases

RLP – **R**eceptor-**l**ike **P**roteins

RNA-Seq – **RNA** **S**equencing

ROS – **R**eactive **O**xygen **S**pecies

T1SS – **T**ype **1** **S**ecretion **S**ystem

T2SS – **T**ype **2** **S**ecretion **S**ystem

T3SS – **T**ype **3** **S**ecretion **S**ystem

T4SS – **T**ype **4** **S**ecretion **S**ystem

TAL – **T**ranscription **A**ctivator -**l**ike

TCSTS – **T**wo-**C**omponent **S**ignal **T**ransduction **S**ystem

UPGMA – **U**nweighted **P**air **G**roup **M**ethod with **A**rithmetic Mean

I. Introduction

THIS PART IS DIVIDED INTO THE FOLLOWING CHAPTERS:

1 Black Rot Disease

1.1 Distribution and importance

Black rot disease, caused by *Xanthomonas campestris* pv. *campestris* (*Xcc*), was first described in the USA in the late 1800's as a disease of cabbage, rutabaga and turnip (Pammel, 1895). Since then the disease has been described in all continents, wherever Brassica production takes place, and is considered the most important bacterial disease affecting these crops (Vicente & Holub, 2013).

Brassicas are vegetables staple for human consumption throughout the globe and their production has been increasing over the last years. According to the Food and Agriculture Organization of the United Nations (FAOSTAT, 2016), cabbage, cauliflower, broccoli and other brassicas production reached 93,7 million tons in 2013 (Figure 1A). 56% of these crops' global production is originated in Asia (Figure 1A), where the gross production value surpassed 14000 million USD (Figure1B). In Portugal, vegetable production between 2011 and 2014 accounted for 7% of the total crop production reaching nearly 1 million tons (Figure 1C). In 2014 cabbage, cauliflower, broccoli and other brassica production reached 268 thousand tons, representing 27% of the total vegetable production (Figure 1C) and reflecting the socio-economic importance of these crops in the country.

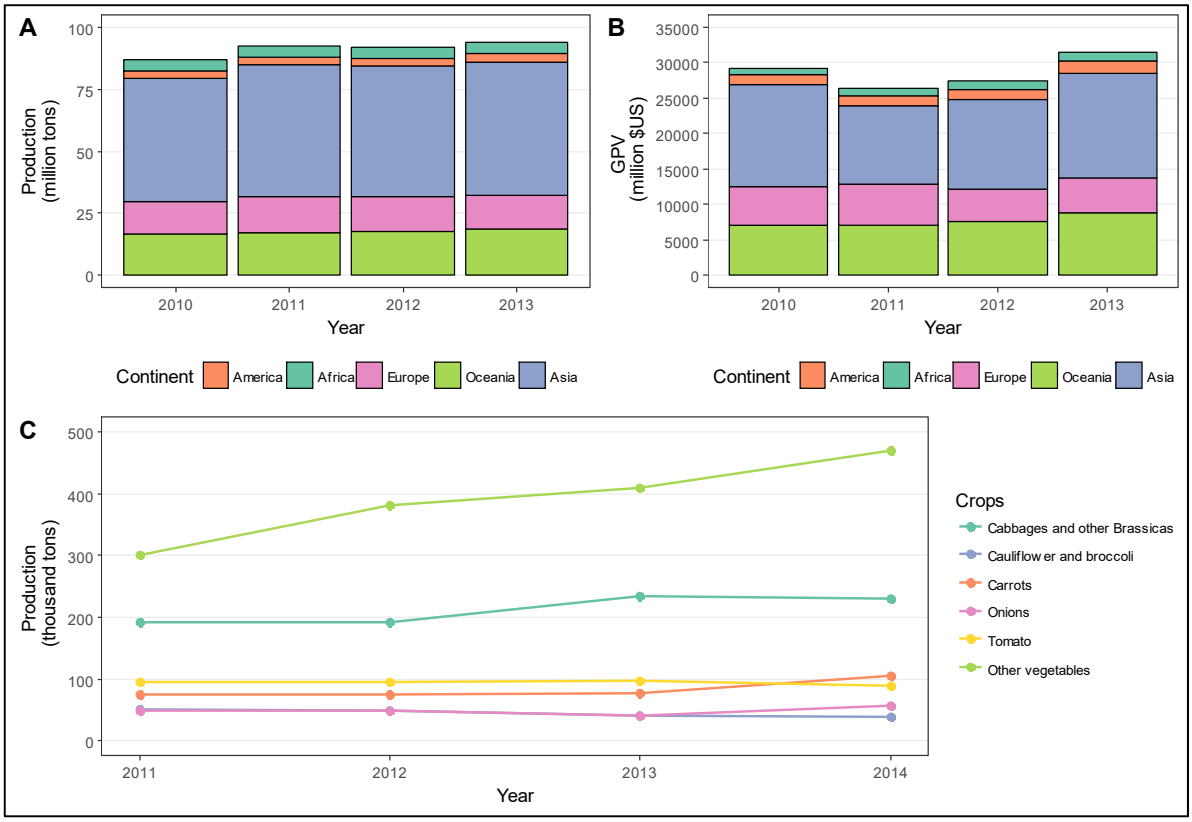


Figure 1 - Global production (A) and gross production value (GPV) (B) of cabbages, cauliflower, broccoli and other Brassicas between 2010 and 2013. Vegetable production (C) in Portugal between 2011 and 2014 (FAOSTAT 2016).

1.1.1 Black rot disease in Portugal

Black rot disease was identified in Portugal, in 1961, affecting cauliflower and savoy cabbage plants that had been imported as seeds from France (Pinto-Ganhão, 1962). Despite its significant impact on Brassica production, characterization of black rot disease and its causative agent *Xanthomonas campestris* pv. *campestris* (*Xcc*) in Portugal has been somewhat incipient.

Between 1994 and 1997, a nation-wide survey was carried out and the disease was found to be present in all Brassica production areas, although disease incidence was higher in fields north of Tagus River margin (Vicente 2004). The author explained this finding as the result of different factors related to the fact that brassicas are traditionally cultivated specially in the northern part of the country, to the higher prevalence of natural inoculum sources and to the use of imported seeds susceptible to the disease.

Additionally, the distinct climatic conditions in the north and the south of the country were also related to disease frequency. While the limiting factor for disease occurrence in the north appears to be the low temperature, in the south the low prevalence of the disease appears to be the low precipitation and relative humidity. In addition to brassicas, spontaneous plants affected by the disease were also identified as evidence of secondary inoculum sources (Vicente, 2004).

The work carried out by Vicente in 2000 included a serologic and pathogenic characterization of the *Xcc* isolates collected between 1994 and 1997 (Vicente, 2000). Protein profiling did not allow differentiating all isolates, and serological characterization revealed that they shared more than 70% similarity. Pathogenic profiling of that set of isolates revealed the presence of races 1, 4 and 6. Race 4 was found to be the most common in Portugal, comprising 53% of the isolates (Vicente, 2000).

More recently, in the context of a polyphasic characterization of a collection of 121 *Xanthomonas* spp. isolates belonging to the Portuguese Collection of Phytopathogenic Bacteria (CPBF - Colecção Portuguesa de Bactérias Fitopatogénicas), the study of 33 *X. campestris* isolates was carried out using a combination of phenotypic, genotypic and phylogenetic approaches (Cruz, 2009). Conservative cultural and biochemical profiles were revealed by the similarity detected between individual biochemical profiles. In contrast, the isolates were characterized by higher genomic diversity, demonstrated by the multiplicity of PCR fingerprinting profiles identified. Phylogenetic analysis of 16SrRNA and *gyrB* gene sequences for eight selected *X. campestris* isolates also suggested a high level of infra-specific diversity. Additionally, biological assays on susceptible hosts revealed a high diversity of pathogenicity profiles, from non-pathogenic isolates to pathogenic isolates showing different virulence patterns. *Xcc* race determination by clipping tests using the standard differential host series revealed the presence of races 1, 4, 6 and 7 as well as atypical race profiles for 22 isolates. While representative the Portuguese *Xcc* population, this set of isolates constitutes a valuable pool of

virulence-associated genes, as well as a raw material for a deeper insight on the underlying molecular mechanisms of this host-pathogen interaction (Cruz, 2009).

1.2 Host Range and symptoms

Black rot disease affects the six most important *Brassica* species: *B. oleracea*, *B. rapa*, *B. nigra*, *B. juncea*, *B. carinata* and *B. napus* (Bradbury, 1986). The genetic and cytogenetic relatedness between these species was established in the 1930's and is often represented in the Triangle of U (U, 1935), as depicted in Figure 2. Additionally, several other Brassicaceae crops, ornamentals and weeds are susceptible to black rot disease, namely radish (*Raphanus sativus*), garden stock (*Mathiolla incana*), white mustard (*Sinapis alba*), field mustard (*Sinapis arvensis*), wallflower (*Erysimum cheiri*) and wild radish (*Raphanus raphanistrum*) (Bradbury, 1986). Some accessions of the model Brassicaceae *Arabidopsis thaliana* are also susceptible to black rot and are systematically used to understand the disease and its causative agent (Simpson & Johnson, 1990).

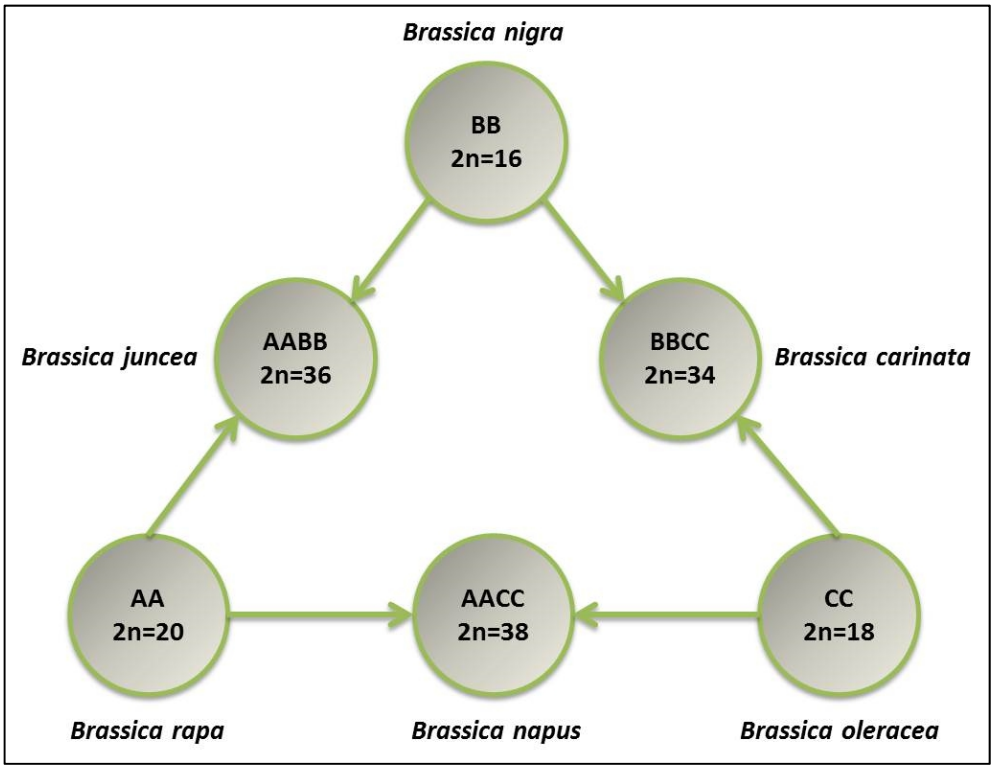


Figure 2 - Genetic relatedness of *Brassica* spp. Three diploid species are shown (*Brassica rapa*, *Brassica nigra* and *Brassica oleracea*), which represent the AA, BB and CC genomes, respectively. Also shown are three tetraploid species (*Brassica carinata*, *Brassica juncea* and *Brassica napus*), which are hybrid combinations of the basic genomes. Diploid chromosome number (2n) is shown. Adapted from U (1935).

Another interesting feature of black rot disease, caused by *Xanthomonas campestris* pv. *campestris* (*Xcc*), is that it can affect *Brassica* spp. plants in any phenological stage, from seed to adult plants (Smith, 1898), as depicted in Figure 3.

Upon germination of infected seeds, cotyledonary leaves exhibit darkened margins and chlorotic lesions surrounding the middle vein. The necrosis of the veins progresses towards the petiole,

leading to the collapse of the cotyledonary leaves, that can remain attached to the plant (Goto, 1990). Plantlets can show stunted growth on one side and may not survive.

Adult infected plants develop typical chlorotic V-shaped lesions with darkened veins on the leaf margins (Goto, 1990), as highlighted in Figure 3. The first sign of the disease is usually a small water lesion on the abaxial surface of the leaf that evolves into a dark spot with a chlorotic halo (Richardson, 1945). In this stage of infection, the dark spots can easily be confused with the leaf spot disease caused by the closely related pathovar *X. campestris* pv. *raphani* (McCulloch, 1929). Upon infection with *Xcc*, the chlorosis progresses towards the middle vein and is accompanied by necrosis of secondary veins, causing the typical V-shaped lesions mentioned earlier (Goto, 1990). When the middle vein is affected, necrosis of the vasculature is evident, leading to dehydration of the surrounding tissues, followed by necrosis and premature abscission of the leaves (Goto, 1990).

Additionally, when the infection becomes systemic, younger leaves can show vascular symptoms and other parts of the plant can be affected, as shown in Figure 3B. The vascular features of the infection also lead to stunted growth and death of young plants. The rapid necrosis of the affected tissues is also promoted by the presence of secondary infections by other plant-pathogens (bacteria or fungi) (Vicente, 2004).

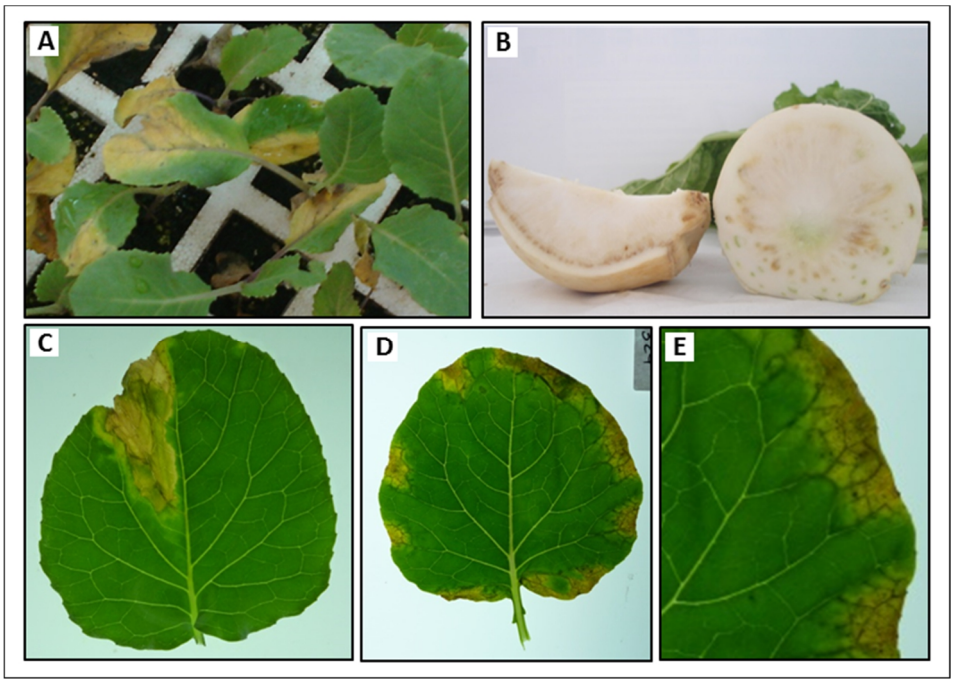


Figure 3 – Black rot disease typical symptoms. (A) *Brassica oleracea* plantlets infected with *Xanthomonas campestris* pv. *campestris*. (B) *Brassica rapa* subsp. *rapa* root showing necrosis of the vasculature. (C) and (D) typical V-shaped lesions in leafs of *B. oleracea*. (E) Detail of (D) highlighting the necrosis of the veins progressing from the leaf margins.

1.3 Disease cycle

1.3.1 Penetration, colonization and dispersal

The black rot disease cycle is described in Figure 4 and can be split down into the three main phases of the infection process: penetration, colonization and dispersal of *Xanthomonas campestris* pv. *campestris* (*Xcc*) bacterial cells within a susceptible host.

The bacterium generally enters the plant through hydathodes, natural openings commonly located on the leaf margins (Smith, 1898). Hydathodes constitute a heterogenic group of structures responsible for the elimination of water, through guttation, when environmental conditions do not favor transpiration, such as high radicular pressure or high relative air humidity (Sinha, 2004). These structures are constituted by two cells that lost the protoplasmic content leaving between them an inactive pore that is permanently open (Sinha, 2004). Beneath this pore there is a parenchymatous tissue (epithem), lacking chloroplasts, that is in direct contact with the last xylemic ramifications of the veins (Sinha, 2004).

Although the bacteria can also enter the host via stomata and cause small necrotic spots, this point of entry is of less importance for disease development, since the pathogen is not able to colonize the surrounding tissues (Cook *et al.*, 1952). Natural phenomena such as wind, rain, insects and other animals, or agricultural practices, such as the use of machinery or irrigation, can cause wounds that serve as additional secondary point of entry for the bacterium (Goto, 1990). Although several studies were carried out, evidence of penetration through the roots has never been found (Vicente & Holub, 2013).

Penetration of *Xcc* through hydathodes occurs by the re-absorption of contaminated droplets of guttation, upon optimal combination of environmental, biological and mechanical factors. Once the bacterium enters the hydathode, it multiplies in the epithem (Meier, 1934). Host cells near bacterial aggregates suffer plasmolysis and cellular organelles collapse, causing the destruction of the middle lamella and collapse of cell walls, forming cavities (Meier, 1934). Bacterial cells multiply in the empty spaces beneath the hydathode finally invading the xylem vessels, allowing dissemination to occur both following and against the water flow (Wallis *et al.*, 1973).

A recent study, based on real time live imaging of *Arabidopsis thaliana* artificially infected with *Xcc*, has shown that bacterial colonization on this host relies on four main steps: 1) movement in the xylem vessel; 2) movement to the next cell; 3) adhesion to plant host cell; and 4) formation of bacterial aggregate (Akimoto-Tomiyama *et al.*, 2014). Those results suggest that the xylem vessel is not only a main colonization site of *Xcc* but also a focal gateway, leading to a systemic infection affecting all parts of the plant, including flowers and seeds (Akimoto-Tomiyama *et al.*, 2014).

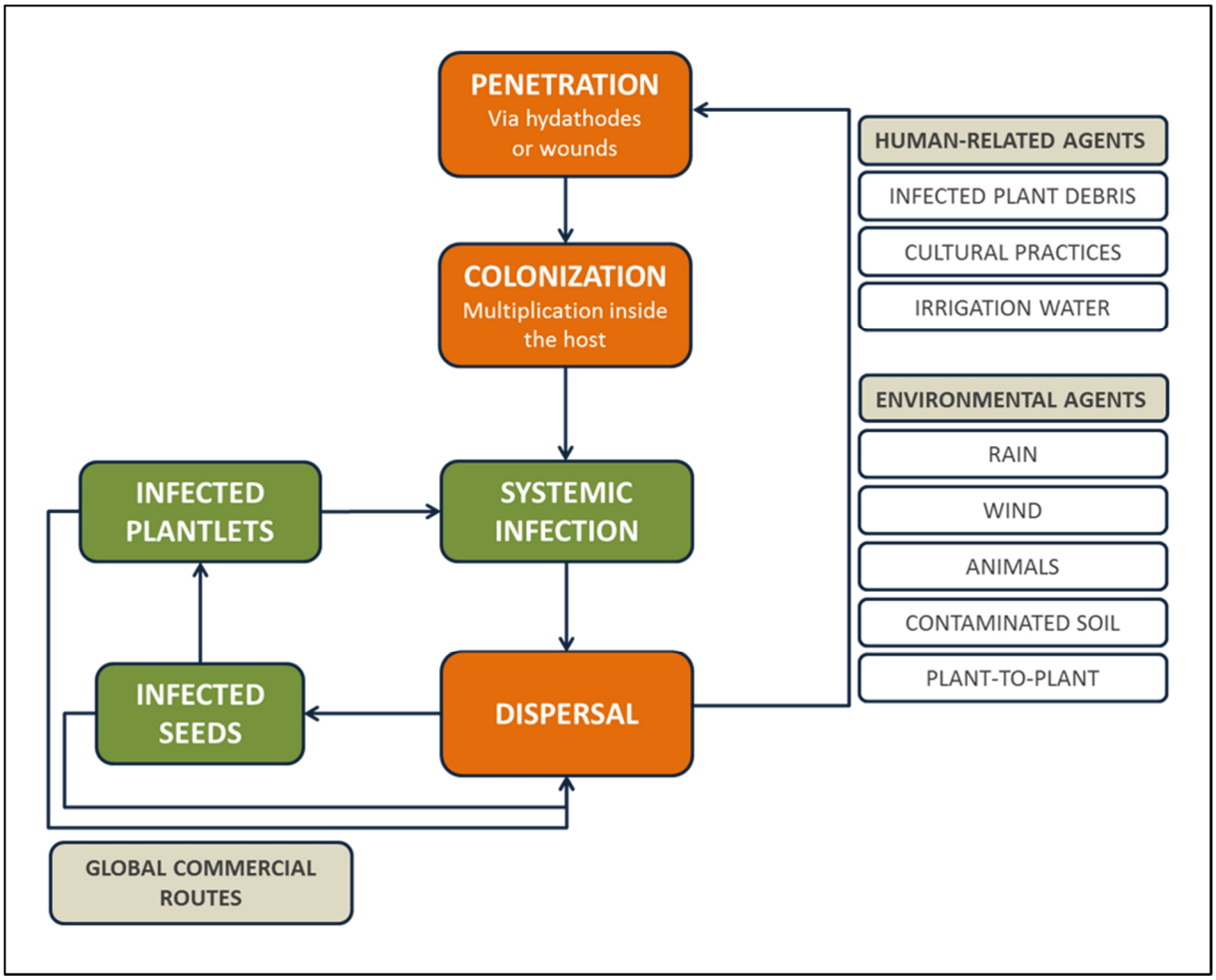


Figure 4 - Black rot disease cycle. Penetration of *Xanthomonas campestris* pv. *campestris* within the host occurs through hydathodes or wounds. The bacteria multiply beneath the hydathodes, colonizing the host. After xylem vessels colonization, the infection becomes systemic and dispersal can occur. Short distance dispersal occurs in the field, via environmental or human-related agents. Long distance dispersal is usually associated with infected seeds and follows global commercial routes.

Long distance dispersal of infected seeds and plantlets follows commercial routes across the globe (Vicente & Holub, 2013). Short distance dispersal of *Xcc* occurs through plant-to-plant contact, mechanical or human manipulation of plants, wind, insects, aerosols and mostly irrigation water or rain. In plant-nurseries producing module-raised transplants, where overhead irrigation is used, bacterial dispersal is especially increased, leading to a higher disease incidence under field conditions (Roberts *et al.*, 2007).

1.3.2 Inoculum sources

Seeds are considered by most authors as the primary source of *Xcc* inoculum. During seed production, *Brassica* seed may become infected with *Xcc* after systemic colonization of plants upon leaf infection, or alternatively, after flower infection (Van Der Wolf & Van Der Zouwen, 2010). Contaminated soil can act as primary inoculum for no longer than two months independently of the host, although *Xcc* survival can be reduced in saturated soils (Dane & Shaw, 1996; Arias *et al.*, 2000). The survival time of the pathogen is increased within plant tissues, particularly in fresh refuse piles,

thus the presence of infected plant debris in the soil can lead to high levels of disease incidence (Kocks, 1996). Non-host plants, especially crucifer weeds (Schaad & Dianese, 1981), can also be sources of inoculum, since *Xcc* can survive as an epiphyte, without causing disease (Dutta *et al.*, 2014). Colonization of non-host seeds through the flower pathway has also been reported in the closely related pathogen *X. campestris* pv. *vesicatoria* (Dutta *et al.*, 2014), thus it is likely that the same could occur in *Xcc*.

1.3.3 Factors affecting disease occurrence

Disease occurrence and progression can be affected by several aspects and is often a tight combination of pathogen, host and environmental factors.

The spatial distribution of several *Xcc* inoculum foci in the field was found to cause higher disease progression, when compared to an infection caused by a single inoculum focus (Kocks *et al.*, 1998). Although no clear correlation has been made for black rot, pathogen virulence has also been identified as conditioning the speed and severity of symptom progression in other vascular plant pathogenic bacteria, namely *Xylella fastidiosa* and *Xanthomonas oryzae* (Yadeta & Thomma, 2013).

Likewise, host cultivar susceptibility plays an important role in disease occurrence and progression, since differences in constitutive natural barriers against pathogens, such as hydathode number or protecting wax production, can result in distinct susceptibilities across cultivars (Freeman & Beattie, 2008). Additionally, disease progression and dispersal from a single inoculum focus was found to be cultivar dependent (Kocks *et al.*, 1999)

Environmental conditions are of topmost importance for disease progression. Since bacterial penetration is dependent of guttation and bacterial multiplication is dependent of warm temperatures, the infection can remain latent when environmental conditions are not conducive. Therefore, in the field, this bacterial disease arises when both temperature and relative humidity are high, and is naturally most severe in tropical, subtropical and humid continental regions (Williams, 1980). Planting season can also influence disease dissemination. Both early and late plantings seem to be less susceptible when compared to intermediate plantings, most likely due to the delayed exposure to high temperatures and precipitation (Vicente, 2004). However, due to the global dispersion of the pathogen and the constraints of climate changes, it is likely for black rot disease to become more significant in northern regions. Moreover, pathogen pollution, defined as the anthropological movement of pathogens outside their natural geographic or host-species range, as well as the intensification, diversification and globalization of agriculture are expected to contribute to plant disease dispersal (Anderson *et al.*, 2004).

1.4 Management and control

Since the 1980's, black rot disease has been considered the most devastating disease affecting *Brassica* spp. production worldwide (Williams, 1980). Due to its easy dispersal and the unavailability of effective chemical/biological treatments, the use of pathogen-free seeds or transplants combined with adequate cultural practices is of utmost importance for the prevention of disease outbreaks. In addition to those approaches and due to the existence of strict regulation limiting the use of antibiotics, especially within the European Union (EU), studies on the biological control of *Xanthomonas campestris* pv. *campestris* (*Xcc*) have gained momentum. However, the selection of resistant plant material is still considered as the ultimate tool in the control of this disease.

1.4.1 Prophylaxis

Like other seed-borne pathogens, *Xcc* is mostly spread at long-distance by the transportation of infected seeds as commodities (Gitaitis & Walcott, 2007). Although apparently straightforward, starting crop production using clean seeds or transplants is not easily achievable.

The practices of avoidance and exclusion by using quarantines and producing seed in semiarid regions to escape infection by bacterial pathogens, has been unable to completely control seed-borne bacterial diseases (Gitaitis & Walcott, 2007). In this manner, testing by seed producers and importing countries is crucial for disease management, thus guaranteeing the use of clean seeds and transplants.

Several international producers test their *Brassica* spp. seed lots following the International Seed Testing Association (ISTA) guidelines to detect *Xcc,* combining bioassays on susceptible hosts and indirect plating onto semi-selective media (Roberts & Koenraadt, 2006; Asma *et al.*, 2014). Despite its effectiveness in detecting the pathogen, seed testing standard thresholds must be adjusted to the type of production and to its destination market, since a contamination level of 0.01% infested seeds has been reported to be sufficient to start an epidemics in transplanted crops (Roberts *et al.*, 2007). Other detection methods for seed-borne *Xcc* have also been described, such as the use of immunofluorescence microscopy (Franken, 1992), PCR with specific primers (Berg *et al.*, 2005; Zaccardelli *et al.*, 2007) and BIO-PCR (Singh *et al.*, 2014).

To prevent outbreaks, several seed treatments are available, including soaking seed lots in hot acidified cupric acetate solution (Schaad *et al.*, 1980), sodium hypochlorite (Sauer, 1986) or calcium hypochlorite (Schultz, 1986) for a suitable amount of time. However, phytotoxicity evidenced by a lower germination rate of treated seed lots is frequent after these treatments. Additionally, most treatments cannot penetrate the seed coat and thus are only effective against external contamination of the seeds (Taylor & Salanenka, 2012). Moreover, seed treatment *per se* is not the solution to the problem, since the bacterium can survive as an epiphyte in secondary inoculum reservoirs, such as spontaneous Brassicaceae plants (Bradbury, 1986).

Careful selection of field crops and nurseries location is also crucial for outbreak prevention. Ideally, *Brassica* spp. crops should be produced in regions where the disease has never been reported, in fields that were never used for *Brassica* production and as far away as possible from other *Brassica* fields (Williams, 1980). Seed production fields must be installed in regions where the climate does not favor disease development, such as semiarid and cooler regions, where environmental conditions are less conducive (Camargo *et al.*, 1995).

Due to the pathogens ability to remain in the soil and to infest other plants, crop rotation, elimination of crop debris and field inspection for the removal of alternative susceptible hosts that can act as a secondary inoculum source are critical to eradicate the disease in the field (Schaad, 1974). Additionally, to avoid plant-to-plant contamination, the plants must be well spaced and excessive humidity must be prevented by keeping the plants well ventilated and avoiding overhead irrigation (Schaad & Alvarez, 1993).

1.4.2 Chemical control

As most bacterial plant diseases, chemical control of black rot in crop open fields and nurseries relies mostly on the use of copper-based products or antibiotics.

The antibacterial activity of copper is well established and the use of copper compounds to treat bacterial plant diseases became standard in the early 1900s (Clark & Stephen, 1905). However, besides causing phytotoxicity, this type of control is also non-selective, and often eliminates only superficial or epiphytic bacteria, thus being highly ineffective against systemic plant pathogens, such as *Xcc* (Mew & Natural, 1993). Additionally, the systematic use of copper compounds has led to the appearance of resistant *Xanthomonas* spp. (Voloudakis *et al.*, 2005), including *Xcc* (Voloudakis *et al.*, 2005; Lugo *et al.*, 2013).

In the 1950s, following the discovery of antibiotics, streptomycin was found to be an excellent chemical tool for the control of several bacterial diseases of plants, with special focus on the control of fire blight caused by *Erwinia amylovora* in the USA (Dunegan *et al.*, 1954; Goodman, 1954). Despite its success, antibiotic resistance in plant pathogenic target bacteria began to appear as early as the 1960s, a few years after the introduction of use of streptomycin (Moller *et al.*, 1981).

The use of antibiotic formulations to control several diseases caused by xanthomonads has also followed the same path. In the 1990s, the use of streptomycin to control bacterial blight of anthurium caused by *X. axonopodis* pv. *diffenbachiae* was followed by the emergence of streptomycin-resistant strains only 2 years after the treatment (Mew & Natural, 1993).

The clinical antibiotic resistance crisis has focused attention on all uses of antibiotics, including in agriculture. It has been suggested that spraying antibiotics across wide extensions of crops in the environment might increase the frequency of antibiotic resistance genes in bacteria living on plant

surfaces and that genes conferring resistance might then be transferred into clinically important bacteria (Stockwell & Duffy, 2012). Although a direct link between antibiotic sprays on plants and antibiotic resistance in clinical bacteria has not been demonstrated, regulations reduce direct human exposure to antibiotics used in plant agriculture.

Currently, only four antibiotics are used for plant disease control, as described in Table 1, and it is unlikely that additional antibiotics will be registered, especially those that are clinically relevant (Stockwell & Duffy, 2012). It is also important to notice that antibiotic use for the control of certain plant bacterial diseases has only been permitted in European Union country members as an exceptional measure, under very specific and restricted circumstances.

Table 1 - Antibiotics authorized for bacterial disease management

Antibiotic	Application method	Target Disease (causative organism)	Countries with authorized used
Streptomycin	Spray	Fire blight (*Erwinia amylovora*)	USA, New Zealand, Canada, Mexico, Germany*, Austria*, Switzerland*
Oxytetracycline	Spray Injection	Fire blight	USA
		Stone fruit bacterial spot (*Xanthomonas arboricola* pv. *pruni*);	USA
		Several vegetable diseases (*Pectobacterium* spp., *Pseudomonas* spp., *Xanthomonas* spp.)	Mexico
Gentamicin	Spray	Fire blight	Mexico
Oxolinic acid	Spray	Fire blight	Israel
		Bacterial panicle blight of rice (*Burkholderia glumae*)	Japan

* in these countries, the use of streptomycin is an emergency measure, subjected to annual review and used under very restricted conditions, authorized by the country national phytosanitary authorities.

1.4.3 Biological control

The inefficacy and negative effects of some of the previously described control methods in addition to the risks of antibiotic use has urged the development of alternative and safer control measures for bacterial plant diseases. Biological control of bacterial plant diseases is usually attained with the external application of specific antagonistic microorganisms that limit the pathogens growth or activities through the production of antimicrobial substances, competition or predation (Pal & Gardener, 2006; Tjamos *et al.*, 2010).

Control of black rot caused by *Xcc* has been successfully reported. The use of antagonistic *Bacillus* spp. and *Pseudomonas* spp. strains was found to reduce disease incidence in some *Brassica* crops (Jalali & Parashar, 1995; Wulff *et al.*, 2002; Massomo *et al.*, 2004; Monteiro *et al.*, 2005). Additionally, the treatment of *Brassica* seeds using a *Bacillus polymyxa* strain also reduced the infection levels, without causing phytotoxicity or negatively affecting the germination rate (Pichard & Thouvenot, 1999).

More broadly, the use of the natural products extracted or fermented from various sources can also be considered as biological control. Although these substances or mixtures may mimic the activities of living organisms, non-living inputs should more properly be referred to as biopesticides or biofertilizers, depending on the primary benefit provided to the host plant (Tjamos *et al.*, 2010).

The use of plant extracts and essential oils to control several xanthomonads has been successfully reported *in vitro* (Bajpai *et al.*, 2011). The extracts and essential oils of several higher plants were found to have antibacterial activity against several strains of *Xcc* (Satish *et al.*, 1999; Deena & Thoppil, 2000; Sokmen *et al.*, 2004; Iacobellis *et al.*, 2005; Bajpai *et al.*, 2010b,a). Despite of the success of these *in vitro* tests, the *in vivo* efficacy of these substances against *Xcc* causing black rot disease has not been demonstrated.

1.4.4 Genetic control

Considering the disadvantages and setbacks of the previously described control methods, the use of resistant cultivars is crucial to the effective control of black rot disease, relying on the identification and characterization of resistance sources (Mew & Natural, 1993). The detection and selection of cultivars with this desired trait is usually attained by screening the susceptibility levels of several host plants to the pathogen, following artificial inoculation under specific environmental conditions (Mew & Natural, 1993).

In a comprehensive study using more than 100 *Brassica* spp. lines, the known races of *Xcc* and the postulated gene-for-gene model between pathogen and host were considered to define types of resistance – race-specific and race-non-specific – and possible homologies of resistance genes (R genes) within and between genomes (Taylor *et al.*, 2002). In this study, resistance to *Xcc* was identified not only in *B. oleracea* (C genome), but also in *B. rapa* (A genome), *B. nigra* (B genome), *B. carinata* (BC genome), *B. juncea* (AB genome), and *B. napus* (AC genome). The relatedness of *B. oleracea* to the remaining five species comprising the triangle of U (Figure 2) and the homologies between their genomes makes them a potential source of resistance to black rot.

However, resistance governed by single R genes is often rapidly overcome by new pathogen races, due to effector diversification (Lannou, 2012). Additionally, it is expected that resistance mechanisms are genetically more complex, based on multiple resistance loci. A total of 18 QTLs with major and minor effects have been mapped on *B. oleracea* chromosomes, suggesting that resistance to black rot disease is complex and quantitatively controlled by multiple genes in *B. oleracea* (Lee *et al.*, 2015).

Race-specific R genes, conferring qualitative resistance, could potentially provide sustainable black rot control, especially if combined with a genetic background of race-nonspecific genes conferring quantitative resistance, like the one frequently observed in several *A. thaliana* accessions

(Vicente & Holub, 2013). Multigenic control of resistance and the emergence of new *Xcc* races overcoming host resistance hinder black rot resistance breeding in *B. oleracea* and no resistance gene has been cloned so far (Vicente & Holub, 2013).

2 *Xanthomonas campestris* pv. *campestris*

2.1 Systematic and taxonomic framework

According to the Bergey's Manual of Systematic Bacteriology (Brenner *et al.*, 2005) and the Taxonomic Outline of Bacteria and Archae (Garrity *et al.*, 2007), the causative agent of black rot disease is the Gram-negative bacterium *Xanthomonas campestris*, included in the following higher taxonomic groups:

Phylum Proteobacteria

Class Gammaproteobacteria

Order Xanthomonadales

Family Xanthomonadaceae

Genus *Xanthomonas*

2.1.1 The genus *Xanthomonas* and its 'type species' *Xanthomonas campestris*

The genus *Xanthomonas* was proposed in the late 1930s, comprising 19 species (Dowson, 1939). Currently, this genus includes 28 species, encompassing pathogens with a very wide host range that can affect both mono and dicotyledonous plants (Euzéby, 1997).

Until the early 1960s, the taxonomy of the genus *Xanthomonas* was mainly based on the host of origin and the phenotypic features, and each xanthomonad displaying different disease symptoms or showing a distinct host range would be described as a new species (Vauterin *et al.*, 1993). This 'new host – new species' concept rapidly led to a complex genus with more than one hundred species. However, the impossibility of distinguishing several species by any feature other than host specificity, i.e. *nomenspecies*, motivated a reclassification based on a drastic reduction of this number to six species – *X. fragariae*, *X. populi*, *X. oryzae*, *X. albilineans*, *X. axonopodis* and *X. campestris* (Dye & Lelliot, 1974). In this reclassification, almost all species were reclassified into *X. campestris*, the type species of the genus. To facilitate the transition into the simpler classification, the names of the former *nomenspecies* were preserved as pathovars of *X. campestris* (Young *et al.*, 1978). Soon, the so-called provisional classification had been around for more than 15 years and *Xanthomonas* became a genus with only a few species but comprising 140 pathovars (Vauterin *et al.*, 1990).

2.2 *Xanthomonas campestris* pv. *campestris* and related pathovars

In the early 1990's, and despite its usefulness for diagnosis purposes, several molecular studies showed the inadequacy of pathovar classification in reflecting the genomic relationships within the genus (Vauterin *et al.*, 1990, 1995). A new classification was then proposed, based on a DNA-DNA

hybridization study performed using more than one hundred *Xanthomonas* strains (Vauterin *et al.*, 1995). As a result, *X. campestris* was emended to include only the six pathovars obtained from crucifers (*Brassicaceae*) – *X. campestris* pv. *campestris*, *X. campestris* pv. *raphani*, pv. *incanae*, *X. campestris* pv. *armoraciae*, *X. campestris* pv. *barbareae* and *X. campestris* pv. *aberrans*. The remaining *X. campestris* pathovars were included as *X. axonopodis* pathovars, reclassified as new species or renamed as *Xanthomonas* spp. (Vauterin *et al.*, 1995).

In 2007, Fargier & Manceau suggested that *X. campestris* should be restricted to only three pathovars – *X. campestris* pv. *campestris* (*Xcc*), *X. campestris* pv. *raphani* (*Xcr*) and *X. campestris* pv. *incanae* (*Xci*) (Fargier & Manceau, 2007). According to these authors, the leaf spot disease thought to be caused by *X. campestris* pv. *barbarea* was not observed in *Barbarea vulgaris*, thus its existence cannot be confirmed. Additionally, there are currently no *X. campestris* pv. *armoraciae* strains in bacterial collections able to cause the leaf spot disease of horseradish described previously (McCulloch, 1929).

The three closely related *X. campestris* pathovars retained are able to cause three distinct diseases – black rot of crucifers, leaf spot of crucifers and bacterial blight of garden stocks, respectively (Fargier & Manceau, 2007).

Xcc, causing black rot disease of crucifers, affects mostly *Brassica oleracea* varieties, the most important Brassicaceous crops, but is also able to infect other *Brassica* crops, radish, ornamental crucifers and related weeds, as well as the model crucifer *Arabidopsis thaliana*.

Xcr is the causative agent of leaf spot of crucifers, affecting not only *Brassica* crops but also *Solanaceae* important crops, such as tomato (*Solanum lycopersicum*) and pepper (*Capsicum anuum*) (Vicente *et al.*, 2006). This disease causes typical necrotic leaf spots surrounded by a chlorotic halo on the leaves and stems of infected plants.

Xci, causing bacterial blight of garden stocks, has been described in *Mathiola* spp. and *Cheiranthus cheiri*. Infected plants show yellowing and wilting of the leaves and the development of necrosis on the stems can also occur (Fargier & Manceau, 2007).

While *Xcc* and *Xci* cause vascular infections, *Xcr* appears to be restricted to the leaf mesophyll, suggesting that, in addition to the different host ranges, the three pathovars of *X. campestris* also display distinct tissue specificities. This tissue tropism has been suggested to be related to the entry point of bacteria in host plants – while *Xcc*, entering through the hydathodes, can cause a vascular disease, *Xcr*, entering through the stomata, is only able to case leaf spots (Cerutti *et al.*, 2017).

2.2.1 Race structure within *Xanthomonas campestris* pv. *campestris*

As for several other plant pathogens, the existence of intrapathovar variation of the host range led to the subdivision of *Xcc* into races, based on differential symptoms on a series of differentiating

cultivars. In 1992, *Xcc* was first divided into five races, on the basis of their pathogenicity against one cultivar of *B. oleracea*, two cultivars of *B. rapa* and one cultivar of *B. juncea* (Kamoun *et al.*, 1992). In 2001, a revised classification including 6 races, three retained from the original structure and three new ones, was proposed (Vicente *et al.*, 2001). Currently, *Xcc* is divided into 9 races (Fargier & Manceau, 2007), according to the distinct reactions of several differentiating *Brassica* lines, as described in Table 2. Although additional cultivars have been proposed to assist in race determination, highlighting the fragilities of this structure (Jensen *et al.*, 2010), they were not tested for all described races and, therefore, were not included in this document. Races 1 and 4 are considered the most predominant worldwide, although race frequency can vary with geographical origin and host of isolation (Vicente & Holub, 2013).

The interaction between plant pathogens and their hosts has often been described using gene-for-gene models based on the concept that plant susceptibility or resistance is governed by matching gene pairs: resistance (R) genes on the host genome and avirulence (A) genes on the pathogen genome (Flor, 1956, 1971). Since most frequently both R and A genes are dominant, resistance occurs whenever one allele of the R gene interacts with the corresponding allele of the A gene, whereas every other combination leads to susceptibility, i.e. disease development (Flor, 1956, 1971). For *Xcc*, two gene-for-gene models have been proposed to provide molecular support for the existence of races. The most accepted model was first proposed in 2001, following the establishment of the six race structure (Vicente *et al.*, 2001) and has undergone further improvement upon the expansion to 9 races in 2007 (Fargier & Manceau, 2007). According to this model, the presence of the nine races described can be explained by the interactions of five gene pairs, R1-R5 and A1-A5 (Table 2).

Table 2 - Race structure of *Xanthomonas campestris* pv. *campestris* and proposed gene-for-gene model. Adapted from Fargier & Manceau (2007).

						Races/ Avirulence genes (A)								
						1	2	3	4	5	6	7	8	9
						A1	…	A1	A1?[a]	…	…	…	A1	A1
						…	A2	…	…	…	…	…	A2	…
						…	A3	A3	…	A3	…	…	A3	A3
						…	…	…	A4	…	…	…	…	A4
Differentiating cultivars	Resistance genes (R)					…	…	…	…	A5	…	A5	…	…
B. oleracea cv. 'Wirosa'	…	…	…	…	…	+[b]	+	+	+	+	+	+	+	+
B. rapa cv. 'Just Right Turnip'	…	…	…	R4	…	+	+	+	-[c]	+	+	+	+	-
B. rapa cv. 'Seven Top Turnip'	…	R2	…	R4	…	+	-	+	-	+	+	+	-	-
B. carinata PI 199947	R1	…	…	R4?[d]	…	-	+	-	-/(+)[e]	+	+	+	-	-
B. juncea cv. 'Florida Broad Leaf Mustard'	R1	…	…	R4?	R5	-	+	-	-	(+)	+	-	-	-
B. oleracea cv. 'Miracle F1'	…	…	R3	…	…	+	-	-	+	-	+	+	-	-

[a] possible presence of avirulence gene A1; [b] susceptibility; [c] resistance; [d] possible presence of resistance gene R4; [e] weakly pathogenic

Another gene-for-gene model, with only four gene pairs (Rc1/avrRc1; Rc2/avrRc2; Rc3/avrRc3; Rp1/avrRp1) has also been proposed to explain the differences in the interactions between 20 *Xcc* strains and 12 *Brassica* cultivars (He *et al.*, 2007b). Although it is expected that resistance genes is this

model may correspond to the R genes in the previously described model, the fact that it relies on *Brassica* spp. lines different from those typically used for differentiating pathovars, poses several limitations to its use (Vicente & Holub, 2013).

Although these gene-for-gene models have been proposed to explain the interaction between bacterial isolates and differential lines, there is currently no solid molecular/genetic supporting data (Vicente *et al.*, 2002; He *et al.*, 2007b).

The recurrent changes to *Xcc* classification at both pathovar and intra-pathovar levels are a direct result of the high level of diversity that characterizes this pathogen, and as new isolates and hosts are identified and studied, variations in its classification are expected.

2.3 Detection and Identification

2.3.1 Morphological, cultural and biochemical features

Xanthomonas campestris is a Gram-negative rod-shaped bacterium ranging from 0.7µm to 1.8µm in length and from 0.4µm to 0.7µm in width, motile by means of a single polar flagellum (Schaad *et al.*, 2001). As with several other xanthomonads, due to the presence of the exopolysaccharide xanthan and of xanthomonadins (photoprotector carotenoids), most *X. campestris* strains form highly mucoid and yellow colonies on YDC agar (Yeast Dextrose Chalk Agar), a medium routinely used for general isolation from plant tissues (Schaad *et al.*, 2001). Additionally, semi-selective media are available for *X. campestris* isolation from seeds, as well as from soil and plant debris (SM agar, FS agar, NSCAA, SX agar) (Schaad *et al.*, 2001).

X. campestris strains are oxidase negative, catalase positive and weak producers of acids from carbohydrates (Schaad *et al.*, 2001). Carbon source metabolic profiling (Hayward & Waterston, 1965), fatty acid analysis (Massomo *et al.*, 2003) and serological tests (Alvarez *et al.*, 1994) have been profusely used in the identification of *Xanthomonas* species. However, these techniques rely on the availability of databases comprising the results obtained for several isolates of different species/pathovars and the clear identification of isolates using these approaches is often hindered by the misidentification of the isolates used as standards in such databases.

2.3.2 Pathogenicity tests

For the identification of both *X. campestris* pathovars and *X. campestris* pv*. campestris* (*Xcc*) races, the inoculation of *Brassica* seedlings is still the most reliable method, although its application is time consuming and of difficult result interpretation, hindering its suitability for routine diagnosis purposes.

Pathogenicity tests should be performed on healthy young plants of very susceptible cultivars, such as *B. oleracea* cv. 'Wirosa F1'. Inoculation methods typically mimic the preferential entry points of the bacteria under natural conditions: through hydathodes or wounds.

Inoculation of plants by clipping the leaf margins with forceps carrying bacterial growth is the most commonly used method. Adding to a more rapid development of symptoms, this method is also easier to perform, more reproducible (since symptoms develop near the inoculation sites) and the plants require less attention (Staub & Williams, 1972; Vicente, 2004).

On the other hand, by spraying the leaves with a bacterial suspension penetration occurs through hydathodes (Staub & Williams, 1972; Robeson *et al.*, 1989). Although symptom development is slower and their location is less predictable, with this type of inoculation, the artifacts of the forceful bacterial entry are avoided. By clipping the leaves, the hosts' physiological features and natural barriers against pathogens are transposed, since the bacteria is directly inserted in the vascular system. In this manner, a less susceptible cultivar (due to lower number of hydathodes or the production of protective waxes) could present the same symptoms as highly susceptible one, leading to biased results (Staub & Williams, 1972).

2.3.3 Molecular methods

Among several molecular techniques, DNA-DNA hybridization is considered the gold standard for classification and delimitation of bacterial species, and it has been successfully used for the clarification of the taxonomy of the genus *Xanthomonas* (Vauterin *et al.*, 1995). However, this approach is experimentally time consuming, particularly when it comes to the identification of large sets of environmental isolates, making it unsuitable for routine identification purposes. Due to these limitations, other molecular based methods were developed.

The emended classification proposed by Vauterin *et al.* in 1995, was mostly confirmed by the use of Repetitive element sequence-based PCR (rep-PCR) and AFLP genomic fingerprinting (Rademaker *et al.*, 2000), although some studies have reported deviations to that reclassification (Schaad *et al.*, 2000; Vauterin *et al.*, 2000). Rep-PCR (using REP, ERIC and BOX primers) has also successfully been used to distinguish *X. campestris* isolates at the species, pathovar and intrapathovar levels (Rademaker *et al.*, 2005; Vicente *et al.*, 2006). Despite of its resolution power, the use of genomic fingerprinting approaches relies on the comparison of gel profiles, which are of difficult standardization between laboratories. Similarly, the use of PCR restricted fragment length polymorphism of *rpfB* and *atpD*, involved in the regulation of pathogenicity factors and in the synthesis of ATP, respectively, can only differentiate *Xanthomonas* species (Simões *et al.*, 2007).

The *hrp* (hypersensitive response and pathogenicity) gene cluster, encoding type III secretion system components, is involved in several pathogenicity mechanisms, and is largely conserved. Thus,

PCR targeting these specific genes of the genome of *X. campestris* has also been used for the identification of this pathogen (Berg *et al.*, 2005). While this approach readily distinguishes *X. campestris* from other xanthomonads and is therefore routinely used for detection of *X. campestris*, it fails to differentiate pathovars within this species (Berg *et al.*, 2005).

Molecular identification of isolates at the species level has also been achieved using methods based on DNA sequencing of constitutive genes allowing the construction of phylogenies with higher resolution power. The study of the 16S rRNA gene (Hauben *et al.*, 1997) and 16S-23S intergenic region (Gonçalves & Rosato, 2002) have also been used to assist in the identification of xanthomonads, but only to the genus level. Although it has been developed to assist in the clarification of the phylogenetic relatedness among strains, the sequencing of *gyrB* gene is a fast and convenient way to identify new isolates (Parkinson *et al.*, 2007, 2009). Even though *gyrB* locus may be sufficiently discriminatory to identify intraspecific diversity at pathovar level, it has not yet been successfully applied to distinguish between isolates within pathovars.

Currently, Multi Locus Sequence Analysis (MLSA) appears to be the most valuable molecular approach. In 2008, a MLSA of the genus *Xanthomonas* successfully distinguished *X. campestris* from other species and was able to detect variations between isolates of the species (Young *et al.*, 2008). In a study comprising 42 isolates from different origins, discrimination of *X. campestris* at pathovar level was achieved using MLSA, confirming the existence of 3 pathovars (Fargier *et al.*, 2011). Although this technique also showed potential to distinguish between isolates at intrapathovar level, there was no clear correlation between the genomic data provided by MLSA and the race structure currently adopted for *Xcc* (Fargier *et al.*, 2011).

The development of whole genome sequencing technology has already been used towards *X. campestris* identification/characterization and it is likely that this tool will be routinely used for the identification and characterization of *Xanthomonas* species, allowing the development of new molecular markers for routine and diagnosis purposes (Vicente & Holub, 2013).

2.4 Genomic features

Comparative analysis of several *X. campestris* has shown that its genome is predicted to encode more than 4000 proteins, in a single circular chromosome ranging from 4.8Mb to 5.3Mb in size, with GC content over 60% (da Silva *et al.*, 2002; Qian *et al.*, 2005; Vorhölter *et al.*, 2008; Bolot *et al.*, 2013a,b, 2015; Desai *et al.*, 2015). Overall, xanthomonads have similar gene contents, although whole-genome comparisons have revealed several inversions, translocations and insertions or deletions.

Xcc type strain ATCC33913^{T}, along with *X. axonopodis* pv. *citri* (*Xac*) strain Xac306 (the causative agent of citrus canker), was the first fully sequenced xanthomonad (da Silva *et al.*, 2002). The

comparison between the genomes of these two plant pathogens, with distinct host ranges, revealed that more than 80% of genes are shared, and that gene order is conserved along most of their respective chromosomes. This study also showed several groups of strain-specific genes that motivated subsequent studies to explain the differing host specificities and pathogenic processes between *Xac* and *Xcc.*

Whole genome sequencing has also been used as a tool to compare other genomic features of several *Xcc* strains, to explain different host specificities and to further assess the molecular mechanisms that control pathogenesis.

In the early 2000s, the spontaneous rifampicin-resistant strain *Xcc* 8004, derived from *Xcc* NCPPB1145, was sequenced and its genome was compared to that of strain ATCC33913^T (Qian *et al.*, 2005). Instead of relying on pinpoint mutations, the genetic variation between those two *Xcc* strains was found to be due to a significant rearrangement across the replication axis between two IS1478 elements and to the loss and acquisition of gene blocks. The presence of these two insertion sequences at each extreme end of the predicted recombination site on the *Xcc* ATCC 33913 genome suggests that this rearrangement may have resulted from a single round of recombination and that they may be the preferred sites for homologous recombination during replication. Using a transposon-based mutagenesis system, the authors also screened for virulence-reduced mutants against the host cabbage *B. oleracea* cv. 'Jingfeng 1' (Qian *et al.*, 2005). Of the resulting 16512 transformants, 75 non-redundant pathogenicity-reduced mutants were identified, allowing the identification of 39 new genes and 11 intergenic regions related to pathogenesis in *Xcc*.

The genome of strain B100, commonly used as a model strain to investigate the biosynthesis of the cell surface polysaccharides, was sequenced and the generated data was used to reconstruct the metabolic pathways involved in xanthan biosynthesis (Vorhölter *et al.*, 2008). Comparative genomics revealed that the genome of strain B100 is highly conserved when compared to that of strain 8004, while differing from that of strain ATCC33913^T by the inversion of a large chromosomal fragment. The higher degree of homology between the genomes of *Xac* 306 and *Xcc* ATCC33913^T when compared to the variation between *Xcc* ATCC33913^T and strains *Xcc* B100 and *Xcc* 8004 is consistent with the interpretation that *Xac* 306 and *Xcc* ATCC 33913 may represent the prototypes of these xanthomonads, while strains *Xcc* 8004 and *Xcc* B100 may have been originated by recent recombination events (Qian *et al.*, 2005).

In the early 2010s, the draft genomes of four *Xcc* strains – one American strain formerly described as *X. campestris* pv. *armoraciae* and three Chinese strains highly pathogenic on *Arabidopsis thaliana* - were released (Bolot *et al.*, 2013a,b).

Additionally, in 2015, four *Xcc* strains - two strains from the predominant races 1 and 4 and two additional strains from New Zealand – were fully sequenced (Bolot *et al.*, 2015; Desai *et al.*, 2015).

Further comparative genomic studies based on these draft genomes are still ongoing and will elucidate the genomic features of these strains.

The most important features of the eleven *Xcc* available genomes are presented in Table 3.

Table 3 - Features of the available *Xanthomonas campestris* pv. *campestris* genomes.

Strain	Isolation			Race	Genome size (bp)[b]	% GC[c]	# CDS[d]	rRNA[e]	tRNA[f]	Reference
	Host	Country	Year							
ATCC33913^T	*B. oleracea* var. *gemmifera*	UK	1957	3	5 076 187	65.00	4 181	2	54	da Silva *et al.* (2002)
8004	*B. oleracea* var. *botrytis*	UK	1958	9	5 148 708	64.90	4 273	2	54	Qian *et al.* (2005)
B100	…[a]	…	…	1	5 079 002	65.00	4 471	2	54	Vorhölter *et al.* (2008)
CN14	*B. juncea* var. *foliosa*	China	2003	…	4 989 674	65.00	4 733	3	53	Bolot *et al.* (2013a)
CN15	*B. rapa* subsp. *chinensis*	China	2003	…	5 019 206	65.00	4 800	4	52	
CN16	*B. rapa* subsp. *pekinensis*	China	2003	…	5 017 785	65.00	4 793	4	52	
Xca5	…	USA	…	1	4 990 056	65.20	4 592	2	51	Bolot *et al.* (2013b)
CFBP1869	*B. oleracea*	Ivory Coast	1976	1	5 008 832	65.20	4 535	3	53	Bolot *et al.* (2015)
CFBP5817	*B. oleracea* var. *botrytis*	Chile	2001	4	4 965 622	65.20	4 432	3	53	
ICMP4013	*B. oleracea* var. *capitata*	New Zealand	1974	…	4 972 211	64.55	4 428	6	55	Desai et al. (2015)
ICMP21080	*B. oleracea*	New Zealand	2008	…	4 927 810	64.60	4 371	6	54	

[a] … information not available; [b] bp, base pairs; [c] GC content; [d] number of coding sequences; [e] ribosomal RNA; [f] transfer RNA

In addition to the comparison of *Xcc* isolates, whole genome sequencing has been used to study the genomic differences between pathovars of *X. campestris* that lead to differing host range as well as distinct tissue specificities.

In 2011, the genome of *X. campestris* pv. *raphani* (*Xcr*) strain 756C was also fully sequenced and compared to other xanthomonads, in an attempt to understand the diversity of host and tissue specificities that characterize the genus *Xanthomonas* (Bogdanove *et al.*, 2011). The results suggested a model in which complex sets of adaptations at the level of gene content account for host specificity, while subtler adaptations at the level of amino acid or noncoding regulatory nucleotide sequence are responsible for tissue specificity.

In 2015, sequencing of the genomes of two *X. campestris* pv. *incanae* strains (CFBP1606R and CFBP2527R), one *Xcr* strain (CFBP5828R) and one non-pathogenic strain formerly described as *X. campestris* pv. *barbareae* (CFBP5825R) revealed that pathovars *Xci* and *Xcr* are expected to have as much genomic diversity as *Xcc* (Roux *et al.*, 2015).

Whole genome sequencing has thus facilitated functional analyses aiming to understand the molecular basis of virulence, host specificity and pathogenesis, through random mutagenesis, signature-tagged mutagenesis and homology analysis (Ryan *et al.*, 2011). Virulence genes and other virulence-associated factors will be further discussed in section 3 of this Introduction.

3 Molecular Mechanisms of Host-Pathogen Interaction

Host pathogen interactions can be identified based on the ability of the pathogen to cause symptoms. While in a compatible interaction the pathogen can escape plant defenses and successfully multiply, in an incompatible interaction the pathogen is recognized by the plant immune system, leading bacterial multiplication to a halt (Jones & Dangl, 2006).

The activation of plant defenses relies on the recognition of invading organisms, which can be achieved by: i) transmembrane pattern recognition receptors (PRRs), responding to slowly evolving pathogen-associated molecular patterns (PAMPs); and ii) intracellular Nucleotide-Binding Leucine-Rich Repeats (NB-LRR) proteins that recognize pathogen effectors, either directly or through their effects on the host targets (Jones & Dangl, 2006).

As depicted in Figure 5, plant innate immunity can thus be represented by four major phases:

1) PAMPs are recognized by PRRs, resulting in PAMP-triggered immunity (PTI), which can stop further colonization (Figure 5A);

2) successful pathogens deploy effectors that contribute to pathogen virulence and can suppress PTI, resulting in effector-triggered susceptibility (ETS) (Figure 5B);

3) a given effector is recognized by a NB-LRR protein resulting in effector-triggered immunity (ETI), which leads to disease resistance and cell death at the infection site by hypersensitive response (HR) (Figure 5C);

4) natural selection drives pathogens to avoid ETI, by shedding the recognized effector (Figure 5D) or by diversification of the effector repertoire, through the acquisition of additional effectors that can suppress ETI. New resistance specificities of the host can induce ETI yet again.

In the following subsections, each phase of the molecular mechanisms governing plant immunity and susceptibility will be covered in detail, especially focusing on *Xanthomonas campestris* pv. *campestris* (*Xcc*) interactions with susceptible hosts.

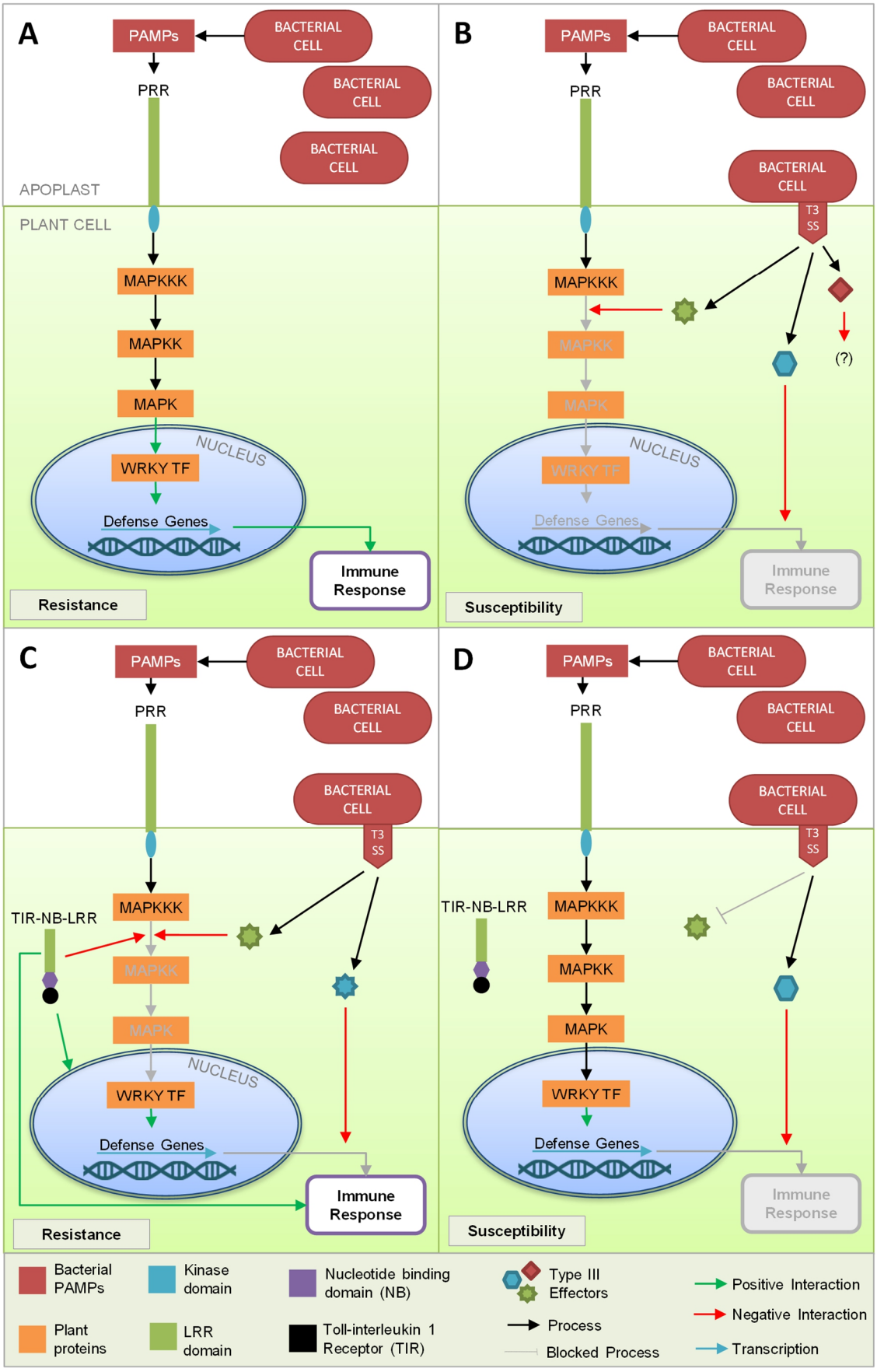


Figure 5 – Molecular mechanisms governing resistance or susceptibility in plant disease. (A) PAMP-triggered immunity (PTI); (B) Effector-triggered susceptibility (ETS); (C) Effector-triggered immunity (ETI); (D) Pathogen evasion of ETI, through shedding of the recognized effector, resulting in susceptibility but less virulence. Adapted from Bent & Mackey (2007).

3.1 PAMP-Triggered Immunity

The first layer of plant immunity is activated upon recognition of pathogen-associated molecular patterns (PAMPs) or endogenous danger-associated molecular patterns (DAMPs), by pattern recognition receptors (PRRs) on the external face of the host cell (Gómez-Gómez & Boller, 2002). Due to the conserved nature of PAMPs across species, genera, family, or class, PAMP- triggered immunity (PTI) has the potential to fend off multiple microbes, pathogenic or not. Thus, PRRs can provide resistance to most non-adapted pathogens, as well as contribute to basal immunity during infection (Zipfel, 2014).

PAMP recognition by PRRs triggers defense responses which include the production of reactive oxygen species (ROS) in a global response commonly known as oxidative burst, changes in the plant cell wall, induction of antimicrobial compounds and the synthesis of pathogenesis-related proteins (Newman *et al.*, 2013). Among the several PAMPs that bacteria shed into the apoplast upon host colonization, flagellin, elongation-factor Tu, peptidoglycan (PGN) and lipopolysaccharides (LPS) play an important role in PTI. The interaction between these PAMPs and their receptors, as well as the following signaling pathways and defense responses are highlighted in Figure 6, and the specificities of known PAMP-PRR interactions in *Xanthomonas campestris* pv. *campestris* and susceptible hosts are further discussed.

Flagellin

Flagellum-dependent motility plays a central role in enabling bacterial colonization of favorable environments and is therefore crucial for the overall pathogenicity of several plant pathogenic bacteria (Newman *et al.*, 2013). The filament of a flagellum is a tubular structure made up of 11 protofilaments, which are nearly longitudinal helical arrays of many hundred 45-kD flagellin molecules (O'Brien & Bennett, 1972). Stray or partially degraded flagellin monomers can leak into the environment during the construction of flagella or as debris associated with bacterial colonies, and be recognized by host cells, inducing defense responses (Newman *et al.*, 2013).

In plants, bacterial flagellin is recognized by the PRR Flagellin-Sensitive 2 (FLS2) through the direct binding of a conserved stretch of 22 amino acid located close to the N terminus of flagellin, the immunogenic epitope flg22 (Gómez-Gómez & Boller, 2002). Binding of flg22 to FLS2 leads to the immediate recruitment of the Leucine Rich Repeat-RK (LRR-RK) Brassinosteroid Insensitive 1 - Associated Kinase 1 (BAK1), which acts as a co-receptor for flg22 and is required for the full activation of FLS2 and flg22-triggered immune signaling (Chinchilla *et al.*, 2007). Formation of this active complex results in transphosphorylation of the respective kinase domains of the PRR and BAK1 (Zipfel, 2014).

Downstream of this active complex signaling is mediated by mitogen-activated protein kinases (MAPK), ultimately leading to the transcriptional regulation of defense genes by WRKY superfamily

transcription factors (WRKY TFs) (Gómez-Gómez & Boller, 2002; Bent & Mackey, 2007; Panstruga *et al.*, 2009). WRKY TF superfamily is comprised by positive and negative regulators containing the typical WRKYGQK peptide sequence and are known to be involved in the regulation of disease resistance networks in several organisms (Pandey & Somssich, 2009).

In addition to flg22, plants also seem able to recognize other epitopes within flagellin. In *Xcc* strains, a single amino-acid polymorphism in flg22 was found responsible for dramatic variations in the ability to elicit FLS2-mediated responses in *Arabidopsis* and in the growth ability of these genetically diverse strains (Sun *et al.*, 2006). Interestingly, isogenic *Xcc* strains expressing eliciting or noneliciting flagellin exhibit similar virulence on *Arabidopsis*, although a preceding flagellin-mediated elicitation of FLS2-dependent responses can constrain their growth (Sun *et al.*, 2006). This complex interaction points to the idea that flagellin detection systems are bypassed by *Xcc*, although host-pathogen interaction is a multilayered system of defense and counterdefense, and other PAMPs can still halt plant colonization by *Xcc* (Sun *et al.*, 2006).

Elongation-factor Tu

As one of the most conserved and abundant protein inside bacterial cells, elongation-factor Tu (EF-Tu) plays a major role in protein biosynthesis, being associated with the ribosome during the elongation phase (Jeppesen *et al.*, 2005). The elicitor activity is attributed to a highly conserved 18aa epitope located on the N-terminus of EF-Tu, named elf18 (Zipfel, 2014).

As depicted in Figure 6, recognition of the conserved EF-Tu by EFR (EF-Tu Receptor) seems restricted to the Brassicaceae family (Kunze, 2004), although transgenic expression of EFR in Solanaceae plants confirmed that EFR confers elf18 perception, revealing that interfamily transfer of plant PRRs can be used to engineer disease resistance in crops (Zipfel *et al.*, 2006).

Like flagellin recognition, perception of EF-Tu can also occur via other epitopes. In fact, treatment with EFa50, a 50aa epitope derived from the central region of EF-Tu induced immune responses in rice (Furukawa *et al.*, 2014).

Despite being independent of the flagellin perception, elf18 recognition by EFR also relies on the recruitment of BAK1 and elicits downstream signaling and response mechanisms similar to those that follow flagellin recognition by FLS2 (Newman *et al.*, 2013; Zipfel, 2014). Indeed, efl18 and fgl22 were found to induce the same gene pool, and the simultaneous recognition of both PAMPs was shown to induce the same MAPKKK pathways without an additive effect (Zipfel *et al.*, 2006).

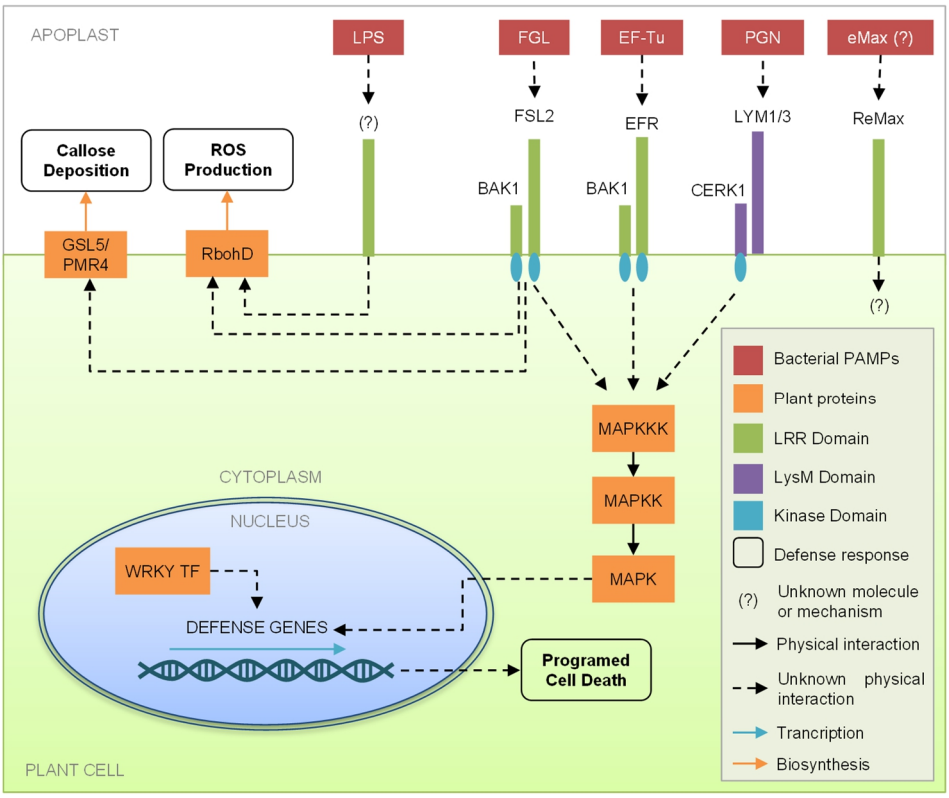


Figure 6 – Overview of PAMP-Triggered Immunity (PTI). Conserved bacterial PAMPs (LPS – lipopolysaccharide; FLG – flagellin, EF-Tu – Elongation factor Tu, PGN – peptidoglycan; eMax – enigmatic MAMP of *Xanthomonas*) are recognized by pattern recognition receptors (PRR) containing Leucin-Rich Repeats (LRR), lysine motif (LysM) and/or kinase domains. Recognition triggers signaling through mitogen-activated protein kinase (MAPK) cascades, activating transcription of defense genes via members of WRKY superfamily (WRKY TF). LPS and FGL also induce reactive oxygen species (ROS) production and cell wall deposition of callose, by activating the expression of RbohD and GSL5/PMR4. Adapted from Panstruga *et al.* (2009) and Zipfel (2014).

Peptidoglycan

Peptidoglycan (PGN) is an essential and unique membrane envelope component of all bacteria, making it an excellent target for the eukaryotic innate immune system (McDonald *et al.*, 2005). Providing rigidity and structure to the cell envelope of both Gram-positive and Gram-negative bacteria, PGN is a complex molecule consisting of numerous glycan chains that are cross-linked by oligopeptides.

PGN perception in plants employs a multimeric receptor system comprising PGN-binding lysine motif receptor-like proteins (LysM-RLPs) and signaling-transducing LysM-Receptor Kinases (LysM-RKs) (Zipfel, 2014). In *Arabidopsis* PGN binds to the LysM-RLPs LYM1 and LYM3, which may form a complex with the LysM-RK CERK1 (Chitin Elicitor Receptor Kinase 1) upon PGN recognition (Zipfel, 2014).

Xcc derived PGN and its constituents functioned as PAMPs in *Arabidopsis* and induced immune responses including the generation of ROS, extracellular pH increase, induced expression of Pathogenesis-Related Protein 1 gene (PR1), and callose deposition (Erbs *et al.*, 2008).

Similarly to what has been described for *Xcc* flagellin perception in *Arabidopsis*, the immunization of tomato with *Staphylococcus aureus* PGN reduced the growth of a subsequent bacterial infection in PGN-treated tissue (Nguyen *et al.*, 2010).

"Orphan" PRRs and bacterial PAMPs

Although several PRR-PAMP pairs have been described, several potential PRRs with ability to recognize bacteria have been identified but their corresponding elicitor PAMPs remain unknown. That is the case of the LRR-RLP ReMAX (Receptor of eMax) which confers recognition of eMax (enigmatic MAMP [Microbial-Associated Molecular Pattern] of *Xanthomonas*) partially purified from *X. axonopodis* pv. *citri* (Jehle *et al.*, 2013).

On the other hand, lipopolysaccharides (LPSs) are known to induce immune responses in plants, but their corresponding PRRs are unknown (Zipfel, 2014).

With multiple roles in plant microbe interactions, LPS is the major component of the outer membrane of Gram-negative bacteria and contributes to the restrictive membrane permeability of these organisms, allowing bacterial growth in unfavorable environments. LPS consists of a lipid (lipid A), a core oligosaccharide, and an O-polysaccharide part (O-antigen). The O-antigen of the LPS from many phytopathogenic bacteria has been shown to consist of oligorhamnans. LPS and its derivatives act as MAMPs and induce innate immune responses in plants (Dow *et al.*, 2000).

In *Xcc*, the purified lipo-oligosaccharides (LOS), LPS without the O-chain, and its derivatives were found to elicit PTI in *Arabidopsis*, inducing a rapid oxidative burst associated with the activation of defense-related genes (Silipo *et al.*, 2005). On the other hand, studies in tobacco cells, showed that neither the lipid A nor the O-chain of the *Xcc* LPS were able to induce the oxidative burst alone; instead, it was found that the inner LPS core part of the LPS molecule was responsible for such early defense response (Braun *et al.*, 2005).

These findings suggest that the activity of LPS and its derivatives as PAMPs can vary substantially across plant hosts. Additionally, pretreatment of plants with LPS (e.g. from heat-killed bacteria) can delay or prevent HR and affect the pattern of gene expression in plants after subsequent inoculation with live bacteria (Sequeira, 1983). The mechanism behind the effect of LPS on HR prevention, termed localized induced resistance (LIR), remains unknown (Newman *et al.*, 2013).

3.2 Effector-Triggered Susceptibility

Successful pathogens can suppress PTI and hence multiply and cause disease. Limitations on the extent of variation (or loss) of specific PAMPs that can be tolerated by bacteria might be a key driving force in the evolution of more mechanisms to suppress PAMP recognition (Abramovitch *et al.*, 2006). This suppression is achieved through the deployment of effectors into the host cell that target basal defense mechanisms, resulting in effector-triggered susceptibility (ETS). Delivery of effectors is achieved mostly via protein secretion systems, although membrane vesicle traffic of virulence-associated proteins into the extracellular milieu has also been reported (Sidhu *et al.*, 2008).

Xanthomonads possess 6 types of protein secretion systems (SS), type I to type VI (T1SS, T2SS, T3SS, T4SS, T5SS and T6SS), differing significantly in composition, function and secreted molecules (Büttner & Bonas, 2010). While all secretion systems play important roles, T2SS and T3SS have been found to play active roles in ETS in *Xanthomonas* spp..

T2SS are responsible for the secretion into the extracellular milieu of plant cell wall degrading enzymes, such as cellulase, polygalacturonase, xylanase and protease (Ryan *et al.*, 2011). There are two identified T2SS in *Xanthomonas* spp. encoded by *xps* and *xcs* gene clusters. Present in all xanthomonads, as well as the related pathogens *Xylella fastidiosa* and *Stenotrophomonas maltophilia*, Xps-T2SS has been shown to contribute directly to virulence in *Xcc*, *X. oryzae pv. oryzae*, *X. oryzae* pv. *oryzicola* and *X. euvesicatoria* (Ryan *et al.*, 2011). Xcs-T2SS is found in some xanthomonads, such as *Xcc*, *X. citri* pv. *citri* and *X. euvesicatoria*, although it does not seem to be directly involved in pathogenesis (Ryan *et al.*, 2011). Although the precise contributions of T2SS to bacterial virulence are unclear, it is assumed that the cell wall degrading activity of their substrates facilitates the assembly of extracellular appendages of virulence-associated T3SS, dedicated to effector translocation into the host cell (Büttner & Bonas, 2010). Contrastingly, T3SS are directly involved in pathogenicity and virulence in most Gram-negative pathogenic bacteria (plant and animal) as a dedicated needle-structure to inject effectors into the host cell, the T3SS.

In plant pathogenic bacteria, the T3SS is encoded by the largely conserved *hrp* (HR and pathogenicity) gene cluster, comprising more than 20 genes arranged in several transcriptional units. Initially described in the bean pathogen *P. syringae* pv. *phaseolicola*, *hrp* genes were found to be essential for pathogenicity and hypersensitive responses in host and non-host plants, respectively (Lindgren *et al.*, 1986). The core T3SS secretion apparatus spans both bacterial membranes and is attached to an extracellular Hrp pilus that transports proteins to host-pathogen interface (Ghosh, 2004). The Hrp pilus is thought to span the plant cell wall and connect to the T3S translocon, a proteinaceous transmembrane channel that inserts into the eukaryotic plasma membrane and mediates the translocation of effector proteins (Ghosh, 2004; Büttner & Bonas, 2010). While most xanthomonads possess the typical Hrp T3SS, *X. albillineans* possesses a non-canonical T3SS similar to the *Salmonella* pathogenicity island 1 (SPI-1) system that is also found in *Erwinia* spp. (Pieretti *et al.*, 2015).

Type III effectors of *Xanthomonas* spp.

The repertoire of type III effectors (T3Es) encoded by individual strains can reach up to 20-30 molecules and varies significantly among closely related strains (Dodds & Rathjen, 2010). This diversification of T3Es appears to be driven by the balance between the need to have enough type III effectors to ensure virulence on susceptible hosts and the need to avoid recognition by plant disease

resistance proteins (Grant *et al.*, 2006). Although mutations of T3SS individual components result in a complete loss of virulence, inactivation of individual T3E genes often does not affect virulence suggesting the existence of functional redundancy among these proteins (Büttner & Bonas, 2010).

Identification of virulence factors, tissue and host-specificity determinants and T3Es has gained impetus with comparative genomics studies. However, due to the lack of conservation in secretion/translocation motifs in T3SP, as well as the lack of mutant phenotypes as a result of functional redundancy, identification and characterization of T3E is often achieved through indirect methods, such as in silico homology searches with known T3Es or eukaryotic signatures (Büttner & Bonas, 2010; Roux *et al.*, 2015).

More recently, transcriptome analysis using RNA-Seq has proven to be a powerful source of information for the structural annotation of T3E candidate genes as well as for the identification of the mechanisms regulating T3E secretion. Very recently, RNA-Seq was used to determine the *hrpG* regulon in *Xcr* strain CFBP5828R (Roux *et al.*, 2015). Along with HrpX, HrpG is a master regulator of *hrp* genes (Wengelnik *et al.*, 1996; Li *et al.*, 2013). The 22 *hrpG*-induced genes unveiled by this RNA-Seq approach constitute a set of T3E candidates that can be further investigated (Roux *et al.*, 2015).

Bacterial effectors have molecular or enzymatic activities that target host proteins and modify their normal cellular functions (Dodds & Rathjen, 2010). Common cellular components and processes manipulated by T3Es include plasma membrane receptors, vesicle trafficking, MAPK signaling and nuclear components (Deslandes & Rivas, 2012). Despite their diversity, T3Es possess a rather small subset of domains and motifs conferring an array of subversive functions within the eukaryotic cell that have been shown to play a role in infection (Dean, 2011).

Two major classes of type III effectors can be found in *Xanthomonas* spp.: transcription activator-like (TAL) effectors and non-TAL effectors. TAL-effectors (also called AvrBs3 effectors) form a particular family of T3Es that act as transcription factors, manipulating the host transcriptome and activating gene expression (Ryan *et al.*, 2011; Hutin *et al.*, 2015; Zhang *et al.*, 2015). Members of TAL family share important structural features, such as nuclear localization signals, an acidic activation domain, and a central region that confers target specificity (Boch & Bonas, 2010). So far, members of the TAL effector family have only been reported in *Xanthomonas* spp. and *Ralstonia* spp. (Zhang *et al.*, 2015). On the other hand, non-TAL effectors are a structurally and functionally diverse group of proteins, found in all xanthomonads. The lack of overlap between the effectomes of *Xanthomonas* spp. and other phytopathogenic bacteria indicates that there are no universal set of effectors to infect plant which suggests that the infection mechanisms vary significantly between pathosystems. However, the existence of functional redundancy among different T3Es suggests that bacteria target conserved plant molecules. The T3Es of *Xanthomonas* spp. are listed in Table 4.

Table 4 – Type III effectors of *Xanthomonas* spp..

Effector	Synonyms	*Xac*[a]	*Xav*[a]	*Xcc*[a]	*Xcr*[a]	*Xci*[a]	*Xoo*[a]	*Xoc*[a]	*Xvm*[a]	*Xvv*[a]
AvrBs1		-	+	XCC2100[b]	-		-	-	-	-
AvrBs2		+	+	XCC0052	-	+	+	+	+	+
AvrBs3 (TAL)	Pth, TAL	+	-	-	-	-	+	+	-	-
XopB		-	+	-	-	+	-	-	+	+
XopC1		-	+	-	-	-	-	-	-	-
XopC2		+	+	-	-	-	+	+	+	+
XopD		-	+	XCC2896	-	-	-	-	-	-
XopD1		-	-	-	-	+	-	-	-	-
XopD2		-	-	-	-	+	-	-	-	-
XopE1	AvrXacE1	+	+	-	-	-	-	-	-	-
XopE2	AvrXacE3, AvrXccE1	+	+	XCC1629	-	+	-	-	-	-
XopE3	AvrXacE2	+	-	-	-	+	-	-	-	-
XopF1	Hpa4	-	+	XCC1218	+	+	+	+	+	+
XopF2		+	+	-	-	-	-	-	-	-
XopG		-	+	XCC3258	-	+	+	-	+	+
XopH	AvrBs1.1	-	+	XCC2099	-	-	-	-	+	+
XopI		+	+	-	-	-	+	+	+	+
XopJ1		-	+	-	-	-	-	-	-	-
XopJ3	AvrRxv	-	+	-	-	-	-	-	+	-
XopJ5	AvrXccB	-	-	XCC3731	-	-	-	-	+	-
XopK		+	+	XCC2899	-	+	+	+	+	+
XopL		+	+	XCC4186	-		+	+	+	+
XopM		-	+	-	-	-	-	-	-	-
XopN		+	+	XCC0231	-	+	+	+	+	+
XopO		-	+	-	-	-	-	+	-	-
XopP		+	+	XCC1247	+	+	+	+	+	+
XopQ		+	+	XCC1072	-	+	+	+	+	+
XopR		+	+	XCC0258	+	+	+	+	+	+
XopS		-	+	-	-	-	-	-	-	-
XopT		-	-	-	-	-	+	-	-	-
XopU		-	-	-	-	-	+	+	-	-
XopV		+	+	-	-	-	+	+	+	+
XopW		-	-	-	-	-	+	+	-	+
XopX		+	+	XCC0529/XCC0530	-	-	+	+	+	+
XopX1		-	-	-	-	+	-	-	-	-
XopX2		-	-	-	-	+	-	-	-	-
XopY		-	-	-	-	-	+	+	+	+
XopZ1		+	+	XCC1975	-	+	+	+	+	+
XopZ2		-	-	-	-	-	-	-	-	-
XopAA	Ecf	-	+	-	-	-	+	+	+	+
XopAB		-	-	-	-	-	+	+	+	+
XopAC	AvrAC	-	-	XCC2565	+	+	-	-	-	-
XopAD		+	+	-	+	+	+	+	+	+
XopAE	HpaF/HpaG	+	+	-	-	+	+	+	+	+
XopAF	AvrXv3	-	-	-	-		-	+	+	+

Effector	Synonyms	*Xac*[a]	*Xav*[a]	*Xcc*[a]	*Xcr*[a]	*Xci*[a]	*Xoo*[a]	*Xoc*[a]	*Xvm*[a]	*Xvv*[a]
XopAG	AvrGf1	-	-	XCC3600	-	+	-	-	+	+
XopAH	AvrXccC	-	-	XCC2109	-	-	-	-	-	-
XopAI		+	-	-	-	-	-	-	-	-
XopAJ	AvrRxo1	-	+	-	-	-	-	+	-	-
XopAK		+	+	-	-	-	-	+	+	+
XopAL1		-	-	XCC1246	+	+	-	-	-	-
XopAL2		-	-	full-length gene downstream XCC3574	-	+	-	-	-	-
XopAM		-	-	XCC1089	-	+	-	-	+	+

[a] *Xac* (*Xanthomonas axonopodis* pv. *citri* 306); *Xav* (*X. euvesicatoria* 85-10); *Xcc* (*X. campestris* pv. *campestris* ATCC33913^T); *Xcr* (*X. campestris* pv. *raphani* 756C); *Xci* (*X. campestris* pv. *incanae* CFBP2527); *Xoo* (*X. oryzae* pv. *oryzae* KACC10331); *Xoc* (*X. oryzae* pv. *oryzicola* BLS256); *Xvm* (*X. campestris* pv. *musacearum* NCPPB4381); *Xvv* (*X. vasicola* pv. *vasculorum* NCPPB702);[b] Locus tag for *X. campestris* pv. *campestris* ATCC33913^T. (http://www.*Xanthomonas*.org/t3e.html).

Although several studies suggested the idea of a core *Xanthomonas* spp. effectome, as more strains are studied and their genomes are sequenced, it becomes clear that T3Es are subjected to variation at genus, species and even pathovar level (Roux *et al.*, 2015). Some effectors appear to be exclusive to *X. campestris*, such as XopAC and XopAL1, while others are totally absent from this species, including XopC2, XopI and XopV.

As seen in Table 4, of the T3Es found in *X. campestris*, five are present in the pathovar type strains for *Xcc* and *Xci,* as well as in *Xcr* strain 756C (widely used for characterization purposes) - XopF1, XopP, XopR, XopAC and XopAL1 – and could therefore be described as core effectome for this species (http://www.*Xanthomonas*.org/t3e.html). However, when screening other strains of *Xcc*, *Xcr* and *Xci*, Roux and coworkers (2015) found that only XopF1, XopP and XopAL1 were present in all strains tested. On the other hand, the T3 secretome of *Xcr* was found to be quite different from that of the remaining pathovars, including XopAD, XopAT and AvrXccA2 which are absent from other pathovars (Roux *et al.*, 2015). Among the same pathovar, differences in the T3 secretome can also be found. For example, there is a mutually exclusive presence/absence of XopD1 and XopD2 among strains of *Xcc* (Guy *et al.*, 2013a). This inter- and intra-pathovar T3E repertoire diversity is yet another reflection of the genomic diversity that characterizes *X. campestris* and hinders the establishment of a core T3 effectome.

Despite the large set of T3Es described in *X. campestris*, the molecular mechanisms supporting leading to ETS remains unclear for most of them. To date, XopAC is the only T3E proven to be an effective suppressor of PAMP-Triggered Immunity, as depicted in Figure 7. In susceptible plants like *A. thaliana*, XopAC uridylylates the BIK1 kinase to inhibit basal resistance and promote virulence (Guy *et al.*, 2013b). However, resistant plants have evolved a decoy substrate which upon uridylylation by AvrAC is recognized by a pseudokinase-immune receptor complex to trigger immunity (Wang *et al.*, 2015).

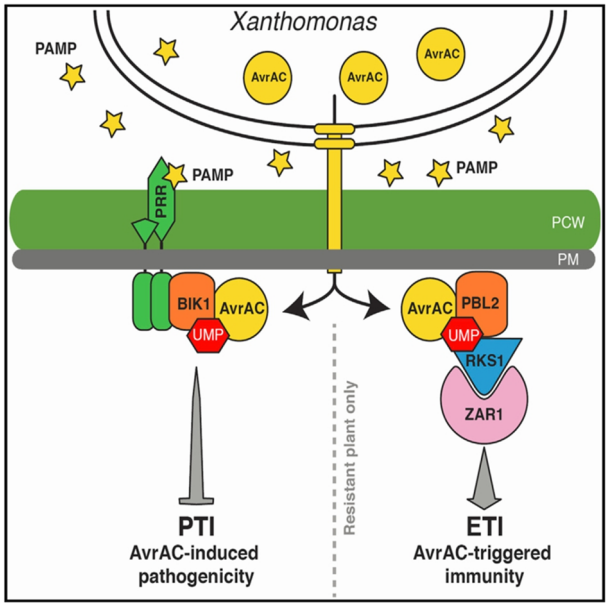


Figure 7 – Model for XopAC (AvrAC) mediated suppression of PAMP-Triggered Immunity (PTI) and Effector-Tiggered Immunity (ETI) in *Arabidopsis thaliana*. *Xanthomonas campestris* pv. *campestris* Type III effector XopAC (AvrAC) inhibits the type VII receptor-like cytoplasmic kinase (RLCK VII) BIK1 (Botrytis Induced Kinase 1), by uridylylation to suppress PAMP -Triggered Immunity.In resistant plants, XopAC(AvrAC) also modifies RLCK VII decoy PBL2 (PBS1-like 2), thus triggering its recruitment to a pre-existing complex made of the RLCK XII RKS1 and the Nucleotide-Binding Leucin-Rich-Repeat ZAR1 causing ETI. (Wang *et al. 2015*)

Regulation of Type III effector genes and other virulence factors in *X. campestris* pv. *campestris*

Secretion of virulence factors is dependent on the correct assembly and functioning of T3SS, coded by *hrp* gene cluster, whose expression is activated when the bacteria enter the plant or are cultivated in certain minimal media (Tang *et al.*, 2006). The expression of *hrp* genes and T3Es in *Xcc* is coordinated by signaling pathways involving several two-component signal transduction systems (TCSTS), and by transcriptional and post-transcriptional regulators (Büttner & Bonas, 2010), as concisely depicted in Figure 8. The detailed roles of each component in this virulence network are further detailed.

TCSTS are broadly used by bacteria, generally consisting of a membrane-bound histidine kinase sensor and a cytoplasmic response regulator (Qian *et al.*, 2008). *Xcc* has more than 50 predicted sensor/regulator pairs, but contribution for virulence has only been demonstrated for a few - RpfC/RpfG, RavA/RavR, ColR/ColS, HrpG/HpaS – which are further described below.

Small diffusible signal molecules (DSFs) produced by bacteria accumulate with the increase of population and can regulate gene expression through corresponding receptor proteins (Wang *et al.*, 2003). In *Xcc* this signal (DSF) has been identified as cis-11-methyl-2-dodecenoic acid, synthesized by the putative enoyl-CoA hydratase RpfF and the fatty acyl-CoA ligase RpfB, both encoded by the *rpf* (regulation of pathogenicity factor) gene cluster (Wang *et al.*, 2003).

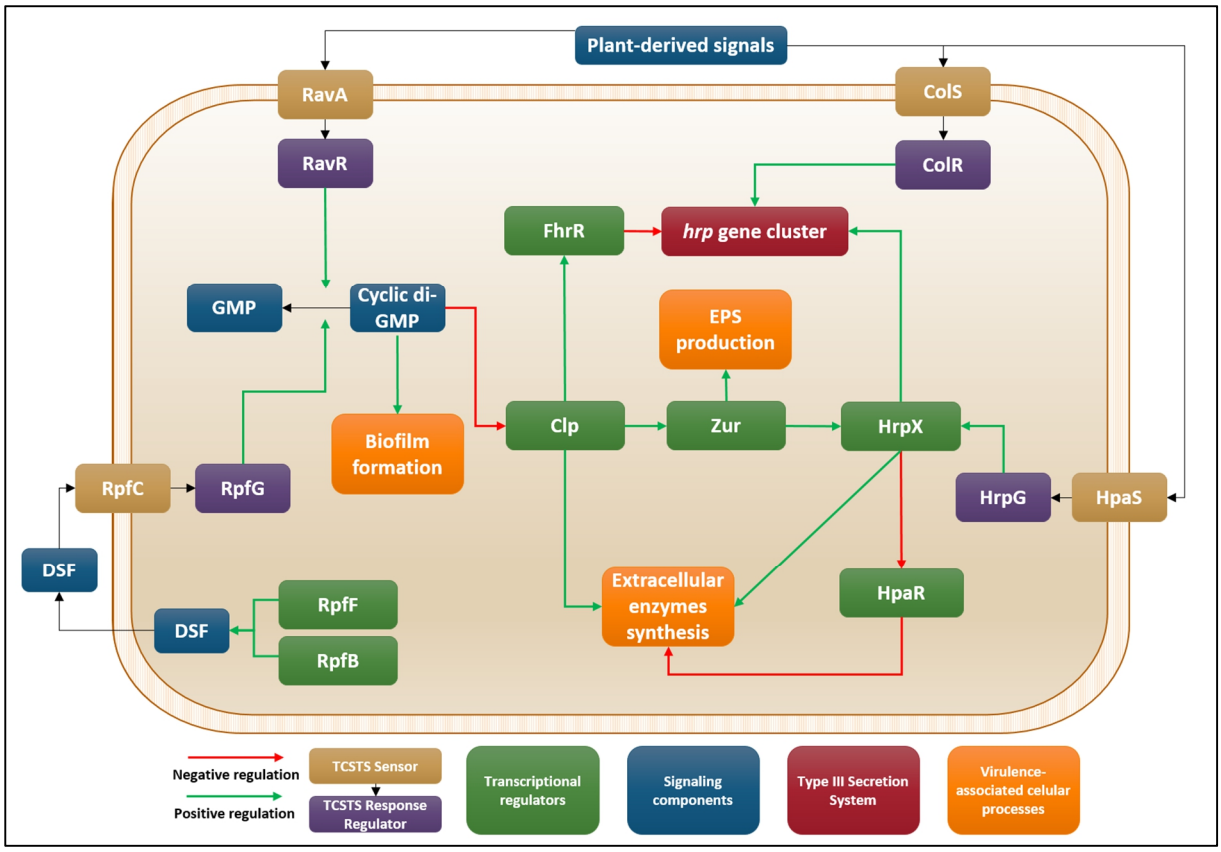


Figure 8 – Model of global regulation of virulence network in *Xanthomonas campestris* pv *campestris*.

Due to its lipophilic nature, DSF is thought to diffuse across the membrane, accumulating in the extracellular milieu in the early stationary phase (Crossman & Dow, 2004). Upon perception of accumulating DSF, the sensor kinase RpfC is autophosphorylated and transfers the phosphoryl group to the receiver domain of the cognate response regulator RpfG (Wang *et al.*, 2003).

RpfG is thought to be involved in the hydrolysis of cyclic di-GMP (bis-(3'-5')-cyclic dimeric guanosine monophosphate), an important messenger molecule (Ryan *et al.*, 2007). Cyclic di-GMP signaling involves proteins with a conserved PilZ domain which have been shown to contribute to virulence in *Xcc* (McCarthy *et al.*, 2008). DSF and RpfC/RpfG were found to activate Clp (CAP[catabolite activator protein]-like protein), a transcriptional regulator containing nucleotide-binding domains that bind promoters of target genes, such as a pectate lyase and an endoglucanase from *Xcc* (Büttner & Bonas, 2010).

Clp also induces the expression of DSF regulon, including genes that encode extracellular enzymes, T2SS and T3SS components and genes involved in EPS synthesis. Among these Clp-regulated genes are two transcription factors - FhrR and Zur (He *et al.*, 2007a). FhrR has been shown to regulate the expression of flagellar, Hrp and ribosomal protein coding genes, whereas Zur is involved in iron uptake, multidrug resistance and detoxification. Additionally, Zur has also been found to induce *hrp* gene expression in *Xcc*, via the transcriptional activator HrpX (Tang *et al.*, 2005). This multilayered regulatory network appears to link virulence and basic, essential cellular functions. Despite these

findings, DSF-mediated control of gene expression differs among strains of *Xcc*, suggest that these mechanisms may be strain-specific (Büttner & Bonas, 2010).

Cyclic di-GMP levels have also been shown to be modulated by the RavA/RavR (Regulation of adaptation and virulence) TCSTS. RavR response regulator promotes EPS production, as well as extracellular protease and cellulase activities (Büttner & Bonas, 2010). Although RavS was formerly predicted to be the histidine kinase (HK) sensor associated with RavR (He *et al.*, 2009), there is a lack of evidence of interaction between both proteins. Additionally, more recent findings suggest that RavA, a predicted HK coded by a gene located downstream of *ravR*, is the actual HK in this TCSTS, regulating RavR phosphorylation levels and enzyme activity (Tao *et al.*, 2014).

hrp gene expression in *Xanthomonas* spp. has also been found to depend on two key transcriptional regulators - HrpX and HrpG (Tang *et al.*, 2006; Büttner & Bonas, 2010). HrpX is an AraC-like transcriptional activator that activates *hrp* operons that often carry a conserved motif termed plant-inducible promoter (PIP)-box (TTCGC-N15-TTCGC) (Büttner & Bonas, 2002). In *Xanthomonas* spp. (and in *Ralstonia* spp.) HrpX is activated by HrpG, an OmpR-type response regulator (Wengelnik *et al.*, 1996). A putative corresponding HrpG sensor kinase, HpaS (*hrp*-associated sensor), was recently discovered (Li *et al.*, 2013). Mutation of *hpaS* in *Xcc* was found to abolish HR and virulence in a similar manner to the effects caused by mutation of *hrpG*. Moreover, HpaS was found to physically interact with HrpG, affecting the expression of *hrpX* (Li *et al.*, 2013).

The TCSTS ColR/ColS was also found to regulate *hrp* gene expression, by inducing *hrpC* and *hrpE* operons, in addition to being involved in control of various cellular processes, including tolerance to environmental stresses (Zhang *et al.*, 2008).

In addition to the regulatory mechanisms described above, virulence gene expression is also a target of post-transcriptional regulators, such as the RNA-binding protein RsmA (Repressor of Secondary Metabolism), a member of a conserved family of proteins initially identified in *Escherichia coli* as repressors of carbon metabolism [carbon storage regulator (CsrA)] (Lapouge *et al.*, 2008). Members of the RsmA/CrsA family bind to the 5' untranslated regions (UTR) of target mRNAs via recognition of sequences containing the conserved trinucleotide motif GGA in the loop structures, thus preventing ribosome binding (Romeo *et al.*, 2013). This translational repression can be relieved by small RNAs that bind to RsmA/CsrA proteins (Lapouge *et al.*, 2008).

In *Xcc*, RsmA has been shown to positively regulate bacterial motility, virulence and HR induction, as well as biofilm formation (Chao *et al.*, 2008). Additionally, RsmA also appears to contribute to enhanced synthesis of extracellular polysaccharides, extracellular endoglucanases and amylases and to the transcript levels of *hrp* and T3 effector genes. More recently, RsmA/CsrA was found to activate T3SS by stabilizing the 5' UTR of *hrpG*, the master regulator of *hrp*/*hrc* genes (Andrade *et al.*, 2014). Because RsmA often acts as a negative translational regulator, positive regulation of gene

expression by RsmA is thought to be indirectly achieved by RsmA-mediated post-transcriptional control of additional regulatory factors, although the identity of potential small RNAs that regulate RsmA activity in *Xanthomonas* spp. remains unclear (Abendroth *et al.*, 2014).

3.3 Effector-Triggered Immunity

Effector proteins were first identified as avirulence (avr) factors, determining host specificity at species, pathovar and race level, by triggering differential responses in certain hosts that possess specific resistance (R) genes (Jones & Dangl, 2006; Ryan *et al.*, 2011).

This one-on-one interaction between avr and R genes activating a signal-transduction cascade to induce plant resistance constitutes the foundation of the successful gene-for-gene model described by Flor in the early 1950's (Flor, 1956).

Intracellular recognition of bacterial effectors is mostly achieved by NLR (nucleotide-binding/leucine-rich-repeat) receptors, thus reinstating and amplifying the PTI basal defense network, in a response known as Effector-Triggered Immunity (ETI) (Dodds & Rathjen, 2010; Cui *et al.*, 2015). For this reason, there is a clear overlap between PTI and ETI processes, evidenced by the similarity of defense responses including the rapid influx of calcium ions, oxidative burst, activation of MAPKs, reprogramming of gene expression, callose deposition and localized cell death (HR) (Tsuda & Katagiri, 2010). The employment of compensatory mechanisms within the defense network and the ability of NB-LRR fine-tuning of responses due to mobility across cell compartments and balancing defense-repressing and defense-activating transcription factors makes ETI much more robust than PTI (Cui *et al.*, 2015). Additionally, local resistance reactions induce longer-lasting systemic immunity (systemic acquired resistance), immunizing pathogen-free tissues against subsequent infections (Cui *et al.*, 2015).

Recognition of bacteria effectors by TIR-NB-LRR receptors usually falls under one of three possible models - direct recognition, guard/decoy or bait models (Dodds & Rathjen, 2010). In direct recognition, the effector physically binds to the NLR receptor, triggering the signaling cascade that induces defense responses. In guard/decoy models the pathogen effector modifies an accessory protein, which may be its virulence target (guard model) or a structural mimic of such a target (decoy model). The NLR receptor then recognizes the modified accessory protein, triggering the defense responses. On the other hand, in bait models the interaction between effector with an accessory protein facilitates the recognition by the NB-LRR receptor.

The ETS/ETI mechanisms described for the mode of action of *Xcc* effector XopAC are based on a decoy model, as evidenced in Figure 7.

Pathogen Evasion of ETI

Despite of these recognition mechanisms, bacterial pathogens have developed the ability to evade ETI through the diversification of effectors, leading to a molecular co-evolutionary arms-race between pathogen and host (Figure 5D). Identification of bacterial effectors is therefore crucial for the identification of corresponding resistance genes, so that they can be targeted for breeding, ideally improving the resistance potential of susceptible plants. Additionally, a deeper understanding of this co-evolution can help predict of how changes in selection parameters will affect the evolution of pathogens, at both microscale and population levels (Dodds 2010).

3.4 Studying molecular host-pathogen interaction - from disease physiology to Next Generation Sequencing

Since the early 1900s, the quest for explaining the development of plant diseases has pushed plant pathologists to use increasingly powerful tools to identify and assess virulence pathogen molecules (Schneider & Collmer, 2010). In the history of molecular plant pathology, the introduction of subsequently more efficient tools established four important research eras - disease physiology (early 1900s - 1980s), single-gene molecular genetics (mid 1980s - 2000), genomics (2000s) and next generation sequencing (NGS) (since 2009).

In addition to the genes involved in basic cellular processes that are particularly important during pathogenesis, virulence-related genes can be characterized as falling under three major categories: 1) True virulence genes, encoding effectors; 2) Virulence-associated genes, directing the deployment of effectors (for example, those involved in the secretion of true effectors); and 3) Virulence lifestyle genes, encoding antimicrobial tolerance factors (for example, those involved in the detoxification of plant ROS) and colonization-promoting factors.

3.4.1 The ‘disease physiology’ era

The identification and validation of virulence factors in the disease physiology era focused mainly on the study of pectolytic cell-wall degrading enzymes, since they are ubiquitously produced by many necrotrophic plant pathogens. Despite describing the effects of these enzymes on the host cell wall, these studies failed to address two of the major important questions: "are these enzymes necessary for virulence?" and "how are these enzymes integrated in the host-pathogen interaction scenario?”. Although a few discoveries were made concerning necrotrophs, the molecular mechanisms underlying biotrophic interactions remained unperceived during this era. The biotrophs' arsenal of avirulence genes, firstly suggested in the 1950s by Flors gene-for-gene model, was only revealed with the advent of molecular genetics.

3.4.2 The 'single-gene molecular genetics' era

Homology-driven reverse genetics was used to test the role of the virulence factors discovered in the previous era. Most of these factors failed to meet the recent molecular Koch's postulates, proposed by Falkow in 1988. These guidelines, created with the intention of promoting the use of more rigorous criteria before declaring a causal relationship between a microorganism and a disease (Falkow, 1988, 2004), were as follows: 1) the studied phenotype must be associated with pathogenic organisms; 2) inactivation of the candidate virulence gene must lead to a measurable loss in virulence; 3) reversion of this mutation should lead to restoration of virulence. Failure to meet the molecular postulates revealed the complexity of the molecular mechanisms governing plant disease.

In fact, forward genetics screens of heterologous expressed genes from related pathogens allowed the discovery of avirulence genes, since their expression conferred avirulence phenotypes to otherwise naturally virulent strains. Screens for reduced virulence of transposon-tagged bacterial mutants allowed the identification of genes encoding protein secretion systems and global regulators, such as the *hrp* gene regulon (Lindgren *et al.*, 1986). The major contribution of this era was the establishment that bacterial virulence is determined by non-redundant secretion systems that deploy large sets of redundant virulence effectors (Schneider & Collmer, 2010).

3.4.3 The 'genomics' era

The genomics era of plant pathology began in 2000, with the sequencing of the genome of a *X. fastidiosa* strain causing citrus variegated chlorosis, a disease of citrus (Simpson *et al.*, 2000). The study of this organism had been previously limited, mostly due to its fastidious nutritional requirements. Although at that time only 47% of the predicted coding sequences were assigned with a putative function, this study yielded a large set of candidate virulence genes, identified on the basis of their homology with known virulence genes from related pathogens, as well as unexpectedly amplified gene classes such as iron uptake systems that could explain the chlorosis caused by the organism (Simpson *et al.*, 2000). As for *X. fastidiosa*, genome sequencing of several other bacterial plant pathogens has generated long lists of candidate virulence genes and virulence regulators, many of which would escape detection due to the weak virulence phenotypes conferred (Schneider & Collmer, 2010).

Most plant pathogenic bacteria available genomes were generated using Sanger-based sequencing and allowed for the exploration of all types of biological processes, including pathogenicity and virulence.

3.4.4 The 'Next Generation Sequencing' era

Although the availability of complete bacterial genomes generated in the 'genomics' era of plant pathology is an unquestionable asset, finding and exploiting genetic diversity and variation is

crucial to grasp the fine-tuning mechanisms that control bacterial pathogenesis. The advent of next generation sequencing (NGS) of both host and pathogen genomes has opened the door for comprehensive studies concerning the identification of effector proteins as well as the characterization of effector expression patterns and the identification, through proteomics and gene homology, of host target proteins (Schneider & Collmer, 2010).

With the goal of understanding increasingly complex phenotypes came the development of platforms generating massive amounts of short-reads, for a fraction of the cost and the time of that of the previously available sequencing technologies (Studholme *et al.*, 2011). However, these short-read NGS approaches also led to higher error rates and often failed to resolve structurally complex regions, an especially important limitation for variant discovery and clinical applications and have led to the more recent development of long-read NGS technologies.

Since the first application of high throughput sequencing to accomplish the *de novo* assembly of the *Pseudomonas syringae* pv. *oryzae* genome (Reinhardt *et al.*, 2009), several research teams have applied NGS strategies targeting both genome and transcriptome to characterize bacterial plant pathogens and their interactions with host plants.

Pre-NGS transcriptomics studies have relied on hybridization of targeted oligonucleotides to particular loci for their sequence specificity: either primers binding to target cDNA in quantitative reverse transcription polymerase chain reaction (qRT-PCR), labeled probes binding to RNA in Northern blotting or hybridization of cDNA to probes on microarray chips (Goodwin *et al.*, 2016). Although these methods are still widely used, bacterial gene expression research is gradually shifting to RNA-Seq, a trend mostly driven by the advantages presented by this technology.

As a probe-free technology, RNA-Seq allows for an annotation-independent detection of transcription, particularly interesting for the study of non-model organisms with unavailable genomes and bacteria with large amounts of genetic variation. In addition to being much more precise, direct mapping to a reference genome also provides an escape to hybridizations' setbacks such as light or radioactivity saturation, allowing both an increased dynamic range and improved sensitivity (Croucher & Thomson, 2010).

RNA-Seq has been used for the analysis of *Xanthomonas* spp. transcriptome, with the goal of understanding the transcriptional responses of these pathogens to different conditions.

In 2013, RNA-Seq-based transcriptome profiling of *Xcc* 8004 grown in MMX medium (minimal medium for *Xanthomonas campestris*) and NYG medium (peptone yeast glycerol medium) determined that *Xcc* responds to both media differently (Liu *et al.*, 2013). The nutrient starvation caused by MMX medium was found to induce the expression of genes involved in amino acid metabolism, transport

systems, atypical condition adaptation and pathogenicity, including T3SS genes. On the other hand, under minimal conditions, many genes involved in chemotaxis and degradation of small molecules were found to be repressed.

More recently, an RNA-Seq-based gene expression analysis of *X. oryzae pv. oryzae* exposed to rice leaf extract revealed dramatic changes in functional categories of genes related to inorganic ion transport and metabolism, and cell motility (Kim *et al.*, 2016). This *in vitro* system produced a genome-wide time-resolved gene expression profile within 1 h of initial rice-*Xoo* interactions, demonstrating the expression order and interaction dependency of virulence related genes (Kim *et al.*, 2016).

In addition to transcriptome profiling, RNA-Seq based approaches have also helped identified potential small non coding RNAs in several plant pathogenic bacteria (Abendroth *et al.*, 2014). An RNA–seq approach of *Xcc* strain 8004 identified three sRNA (sRNA-Xcc15, -Xcc16, and -Xcc28) whose expression was proven to be rpf-dependent (An *et al.*, 2013). Interestingly, the virulence of a *Xcc* triple mutant deleted in the three sRNA genes was reduced compared with the wild-type, whereas virulence of single mutants was not altered, suggesting they may play a cumulative role in pathogenesis.

To date, most RNA-Seq studies concerning plant-pathogenic bacteria have been conducted only on cultured strains. However, to fully understand the complexity of the infection process, profiling both bacterial and plant gene expression during pathogenesis is crucial. Given the increased sensitivity and sequencing depth that is currently available, RNA-Seq appears promising for profiling *in planta* gene expression. Although the feasibility of this approach is challenged by limiting factors, such as the relative abundance of individual transcripts and the structural differences between eukaryotic and prokaryotic mRNAs, dual RNA-Seq of plant pathogens is now evolving (Westermann *et al.*, 2012).

4 Research strategy and presentation of results

As highlighted in the Introduction, corresponding to Part I of this thesis, black rot disease caused by *Xanthomonas campestris* pv. *campestris* is still a world-wide devastating disease without effective control measures and, despite the growing body of work on this subject, no resistance genes have been successfully cloned in Brassicaceae crops. To surpass this challenge, research on the molecular mechanisms of host-pathogen interaction is of paramount importance. These mechanisms have been mostly inferred from an *in vitro* stand-point in conditions that aim to mimic the host environment, resulting in a lack of information concerning the pathogen *in planta*. In this context, and to fulfill that knowledge gap, a line of research has been developed, in a collaboration between the Phytobacteriology Laboratory of INIAV and the Microbiology & Biotechnology and Gene Expression and Regulation Groups of BioISI.

The work presented in this thesis is the follow-up and further development of a preliminary study completed in 2009, in which a collection of more than 120 Portuguese *Xanthomonas* spp. isolates was globally characterized using an integrative approach combining phenotypic, genotypic and phylogenetic data. Several interesting features of the isolates belonging to *X. campestris* were then identified, namely their high degree of genetic diversity and pathogenic profiles, when compared to isolates of from other geographic origins. Along with the fact that Portugal is long known as an important center of domestication for *Brassica oleracea* crops, those findings led to the conception that the Portuguese *X. campestris* strains constitute a pristine genetic diversity reservoir, making them ideal candidates for the study of host/pathogen interactions.

Considering that black rot of crucifers is still a devastating disease worldwide without effective control measures and that Portuguese strains were already found to be an unique sub-population for this pathogen, the work presented in this thesis was outlined to fulfil two major objectives: i) to characterize a set of *Xanthomonas campestris* strains collected in Portugal in terms of pathogenicity, virulence, population structure and phylogenetic diversity; and ii) to determine the *in planta* transcriptomic profile of differentially virulent strains of *X. campestris* pv. *campestris* during the infection process. On the other hand, since *Xcc* is considered a model organism affecting the model plant *Arabidopsis thaliana*, this work was also expected to contribute for the fundamental knowledge of host-pathogen interactions, not only in black rot, but also in other bacterial plant diseases.

In such framework, the results from the experimental work are presented in two main parts, corresponding to Part II and III of this thesis.

In Part II, a characterization of a set of *X. campestris* isolates collected in Portugal is described, following a further subdivision into two chapters. In Chapter 1 the phenotypic diversity of that set of

isolates is explored, by using pathogenicity tests to accurately determine *X. campestris* pathovar and *X. campestris* pv. *campestris* races, and PCR screening of known virulence genes to allow the establishment of virulence profiles. Additionally, a phylogenetic analysis using the sequence of the constitutive *gyrB* gene was carried out, to clarify the phylogenetic relationships characterizing the Portuguese population of *X. campestris*. In Chapter 2, the genetic diversity and population structure of a global set of *X. campestris* pv. *campestris* strains is further inferred, using a four loci Multi-Locus Sequence typing scheme.

Part III of this thesis builds on Part II, and is further subdivided into two additional chapters. In Chapter 1, a comprehensive virulence assessment of the Portuguese *X. campestris* pv. *campestris* population is reported. After inoculation in three susceptible hosts, the strains representing the extremes of the virulence spectrum were selected and used in the study carried out in Chapter 2. In this chapter, using an *in planta* RNA-Seq approach, the genes that were differentially expressed by each strain during the infection process were identified and their contribution for pathogenicity and virulence is further discussed.

In Part IV, the main conclusions and contributions of the work developed are presented, along with related ongoing research and perspectives for future research.

References

Abendroth U, Schmidtke C, Bonas U, 2014. Small non-coding RNAs in plant-pathogenic *Xanthomonas* spp. *RNA Biology* **11**, 457–463.

Abramovitch RB, Anderson JC, Martin GB, 2006. Bacterial elicitation and evasion of plant innate immunity. *Nature reviews. Molecular cell biology* **7**, 601–611.

Akimoto-Tomiyama C, Furutani A, Ochiai H, 2014. Real time live imaging of phytopathogenic bacteria *Xanthomonas campestris* pv. *campestris* MAFF106712 in "plant sweet home". *PloS one* **9**, e94386.

Alvarez AM, Benedict AA, Mizumoto CY, Hunter JE, Gabriel DW, 1994. Serological, Pathological, and Genetic Diversity Among Strains of *Xanthomonas campestris* Infecting Crucifers. *Phytopathology* **84**, 1449–1457.

An S-Q, Febrer M, McCarthy Y *et al.*, 2013. High-resolution transcriptional analysis of the regulatory influence of cell-to-cell signalling reveals novel genes that contribute to *Xanthomonas* phytopathogenesis. *Molecular microbiology* **88**, 1058–1069.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P, 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in ecology & evolution* **19**, 535–544.

Andrade MO, Farah CS, Wang N, 2014. The post-transcriptional regulator rsmA/csrA activates T3SS by stabilizing the 5' UTR of *hrpG*, the master regulator of *hrp*/*hrc* genes, in *Xanthomonas*. *PLoS pathogens* **10**, e1003945.

Arias RS, Nelson SC, Alvarez AM, 2000. Effect of soil–matric potential and phylloplanes of rotation-crops on the survival of a bioluminescent *Xanthomonas campestris* pv. *campestris*. *European Journal of Plant Pathology* **106**, 109–116.

Asma M, Koenraadt HMS, Politikou A, 2014. Detection of *Xanthomonas campestris* pv. *campestris* on *Brassica* spp . disinfested / disinfected seed with grinding. In: *International Rules for Seed Testing*. 1–14.

Bajpai VK, Cho MJ, Kang SC, 2010a. Control of Plant Pathogenic Bacteria of *Xanthomonas* spp. by the Essential Oil and Extracts of *Metasequoia glyptostroboides* Miki ex Hu *In vitro* and *In vivo*. *Journal of Phytopathology* **158**, 479–486.

Bajpai VK, Dung NT, Suh H-J, Kang SC, 2010b. Antibacterial Activity of Essential Oil and Extracts of Cleistocalyx operculatus Buds Against the Bacteria of *Xanthomonas* spp. *Journal of the American*

Oil Chemists' Society **87**, 1341–1349.

Bajpai VK, Kang S-R, Xu H, Lee S-G, Baek K-H, Kang S-C, 2011. Potential Roles of Essential Oils on Controlling Plant Pathogenic Bacteria *Xanthomonas* Species: A Review. *The Plant Pathology Journal* **27**, 207–224.

Bent AF, Mackey D, 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology* **45**, 399–436.

Berg T, Tesoriero L, Hailstones DL, 2005. PCR-based detection of *Xanthomonas campestris* pathovars in Brassica seed. *Plant Pathology* **54**, 416–427.

Boch J, Bonas U, 2010. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annual Review of Phytopathology* **48**, 419–436.

Bogdanove AJ, Koebnik R, Lu H *et al.*, 2011. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic Xanthomonas spp. *Journal of bacteriology* **193**, 5450–5464.

Bolot S, Cerutti A, Carrère S *et al.*, 2015. Genome Sequences of the Race 1 and Race 4 *Xanthomonas campestris* pv. *campestris* Strains CFBP 1869 and CFBP 5817. *Genome announcements* **3**, e01023-15.

Bolot S, Guy E, Carrere S, Barbe V, Arlat M, Noël LD, 2013a. Genome Sequence of *Xanthomonas campestris* pv. *campestris* Strain Xca5. *Genome announcements* **1**, e00032-12.

Bolot S, Roux B, Carrere S *et al.*, 2013b. Genome sequences of three atypical *Xanthomonas campestris* pv. *campestris* strains, CN14, CN15, and CN16. *Genome announcements* **1**, e00465-13.

Bradbury JF, 1986. *Guide to Plant Pathogenic Bacteria*. Farnham Royal, Slough: CAB International.

Braun SG, Meyer A, Holst O, Pühler A, Niehaus K, 2005. Characterization of the *Xanthomonas campestris* pv. *campestris* lipopolysaccharide substructures essential for elicitation of an oxidative burst in tobacco cells. *Molecular plant-microbe interactions : MPMI* **18**, 674–681.

Brenner DJ, Krieg NR, Staley JT *et al.* (Eds.), 2005. *Bergey's Manual® of Systematic Bacteriology*. Boston, MA: Springer US.

Büttner D, Bonas U, 2002. Getting across--bacterial type III effector proteins on their way to the plant cell. *The EMBO journal* **21**, 5313–5322.

Büttner D, Bonas U, 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* **34**, 107–133.

Camargo LE a, Williams PH, Osborn TC, 1995. Mapping of quantitative trait *loci* controlling resistance of *Brassica oleracea* to *Xanthomonas campestris* pv. *campestris* in the field and greenhouse. *Phytopathology* **85**, 1296–1300.

Cerutti A, Jauneau A, Auriac M-C *et al.*, 2017. Immunity at cauliflower hydathodes controls infection by *Xanthomonas campestris* pv. *campestris*. *Plant Physiology* **174**, 700–716.

Chao N-X, Wei K, Chen Q *et al.*, 2008. The rsmA-like gene rsmA(Xcc) of *Xanthomonas campestris* pv. *campestris* is involved in the control of various cellular processes, including pathogenesis. *Molecular plant-microbe interactions : MPMI* **21**, 411–423.

Chinchilla D, Zipfel C, Robatzek S *et al.*, 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497–500.

Clark H, Stephen D, 1905. On the Bactericidal Action of Copper. *Public health papers and reports* **31**, 175–204.

Cook AA, Walker JC, Larson RH, 1952. Studies on the Disease Cycle of Black Rot of Crucifers. *Phytopathology* **42**, 162–167.

Crossman L, Dow JM, 2004. Biofilm formation and dispersal in *Xanthomonas campestris*. *Microbes and infection* **6**, 623–629.

Croucher NJ, Thomson NR, 2010. Studying bacterial transcriptomes using RNA-seq. *Current opinion in microbiology* **13**, 619–624.

Cruz J, 2009. Diversidade genómica, patogenicidade e sensibilidade a compostos bioactivos em *Xanthomonas campestris* . Faculdade de Ciências da Universidade de Lisboa.

Cui H, Tsuda K, Parker JE, 2015. Effector-triggered immunity: from pathogen perception to robust defense. *Annual review of plant biology* **66**, 487–511.

Dane F, Shaw JJ, 1996. Survival and persistence of bioluminescent *Xanthomonas campestris* pv. *campestris* on host and non-host plants in the field environment. *Journal of Applied Bacteriology* **80**, 73–80.

Dean P, 2011. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS microbiology reviews* **35**, 1100–1025.

Deena MJ, Thoppil JE, 2000. Antimicrobial activity of the essential oil of *Lantana camara*. *Fitoterapia* **71**, 453–455.

Desai D, Li J-H, van Zijll de Jong E *et al.*, 2015. Draft Genome Sequences of Two New Zealand

Xanthomonas campestris pv. *campestris* Isolates, ICMP 4013 and ICMP 21080. *Genome announcements* **3**, e01247-15.

Deslandes L, Rivas S, 2012. Catch me if you can: Bacterial effectors and plant targets. *Trends in Plant Science* **17**, 644–655.

Dodds PN, Rathjen JP, 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature reviews. Genetics* **11**, 539–548.

Dow M, Newman M, Roepenack E Von, 2000. The induction and modulation of plant defense responses by bacterial lipopolysaccharides. *Annual review of microbiology* **38**, 241–261.

Dowson JW, 1939. On the systematic position and generic names of the Gram-negative bacterial plant pathogens. *Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene* **2**, 177–193.

Dunegan JC, Kienholz JR, Wilson RA, Morris WT, 1954. Control of pear blight by a streptomycin–terramycin mixture. *Plant Disease Reporter* **38**, 666–669.

Dutta B, Gitaitis R, Smith S, Langston D, 2014. Interactions of seedborne bacterial pathogens with host and non-host plants in relation to seed infestation and seedling transmission. *PloS one* **9**, e99215.

Dye DW, Lelliot RA, 1974. Genus 11. *Xanthomonas*. In: R. E. Buchanan and N. E. Gibbons, ed. *Bergey's manual of determinative bacteriology*. Baltimore, 243–249.

Erbs G, Silipo A, Aslam S *et al.*, 2008. Peptidoglycan and muropeptides from pathogens *Agrobacterium* and *Xanthomonas* elicit plant innate immunity: structure and activity. *Chemistry & biology* **15**, 438–448.

Euzéby JP, 1997. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *International journal of systematic and evolutionary bacteriology* **47**, 590–592.

Falkow S, 1988. Molecular Koch's postulates applied to microbial pathogenicity. *Reviews of infectious diseases* **10**, 274–276.

Falkow S, 2004. Molecular Koch's postulates applied to bacterial pathogenicity — a personal recollection 15 years later. *Nature Reviews Microbiology* **2**, 67–72.

Fargier E, Fischer-Le Saux M, Manceau C, 2011. A multilocus sequence analysis of *Xanthomonas campestris* reveals a complex structure within crucifer-attacking pathovars of this species. *Systematic and applied microbiology* **34**, 156–165.

Fargier E, Manceau C, 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into

three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**, 805–818.

Flor HH, 1956. The Complementary Genic Systems in Flax and Flax Rust. *Advances in Genetics* **8**, 29–54.

Flor HH, 1971. Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology* **9**, 275–296.

Franken AAJM, 1992. Application of polyclonal and monoclonal antibodies for the detection of *Xanthomonas campestris* pv. *campestris* in crucifer seeds using immunofluorescence microscopy. *Netherlands Journal of Plant Pathology* **98**, 95–106.

Freeman BC, Beattie GA, 2008. An Overview of Plant Defenses against pathogens and herbivores. *The Plant Health Instructor*.

Furukawa T, Inagaki H, Takai R, Hirai H, Che F-S, 2014. Two Distinct EF-Tu Epitopes Induce Immune Responses in Rice and *Arabidopsis*. *Molecular Plant-Microbe Interactions* **27**, 113–124.

Garrity GM, Lilburn TG, Cole JR, Harrison SH, Euzéby J, Tindall BJ, 2007. Taxonomic Outline of the Bacteria and Archaea , Release 7 . 7 March 6 , 2007 . Part 5 – The Bacteria : Phylum " Proteobacteria ", Class Gammaproteobacteria. In: *Taxonomic Outline of the Bacteria and Archaea , Release 7 . 7 March 6 , 2007*. 148–245.

Ghosh P, 2004. Process of Protein Transport by the Type III Secretion System. *Microbiology and molecular biology reviews* **68**, 771–795.

Gitaitis R, Walcott R, 2007. The epidemiology and management of seedborne bacterial diseases. *Annual review of microbiology* **45**, 371–397.

Gómez-Gómez L, Boller T, 2002. Flagellin perception: a paradigm for innate immunity. *Trends in Plant Science* **7**, 251–256.

Gonçalves ER, Rosato YB, 2002. Phylogenetic analysis of *Xanthomonas* species based upon 16S-23S rDNA intergenic spacer sequences. *International journal of systematic and evolutionary microbiology* **52**, 355–361.

Goodman RN, 1954. Fireblight control with sprays of Agri-Mycin, a streptomycin-terramycin combination. *Plant Disease Reporter* **38**, 874–878.

Goodwin S, McPherson JD, McCombie WR, 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333–351.

Goto M, 1990. *Fundamentals of Bacterial Plant Pathology* (I Academic Press, Ed,). San Diego.

Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL, 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annual review of microbiology* **60**, 425–449.

Guy E, Genissel A, Hajri A *et al.*, 2013a. Natural genetic variation of *Xanthomonas campestris* pv. *campestris* pathogenicity on *Arabidopsis* revealed by association and reverse genetics. *mBio* **4**, e00538-12.

Guy E, Lautier M, Chabannes M *et al.*, 2013b. xopAC-triggered Immunity against *Xanthomonas* Depends on *Arabidopsis* Receptor-Like Cytoplasmic Kinase Genes PBL2 and RIPK (C-H Yang, Ed,). *PLoS ONE* **8**, e73469.

Hauben L, Vauterin L, Swings J, Moore ER, 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *International journal of systematic bacteriology* **47**, 328–335.

Hayward AC, Waterston JM, 1965. Xanthomonas campestris C.M.I. *Descriptions of Pathogenic Fungi and Bacteria*, 47.

He Y-W, Boon C, Zhou L, Zhang L-H, 2009. Co-regulation of Xanthomonas campestris virulence by quorum sensing and a novel two-component regulatory system RavS/RavR. *Molecular microbiology* **71**, 1464–1476.

He Y-W, Ng AY-J, Xu M *et al.*, 2007a. *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Molecular Microbiology* **64**, 281–292.

He Y-Q, Zhang L, Jiang B-L *et al.*, 2007b. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome biology* **8**, R218.

Hutin M, Pérez-Quintero AL, Lopez C, Szurek B, 2015. MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Frontiers in plant science* **6**, 1–12.

Iacobellis NS, Lo Cantore P, Capasso F, Senatore F, 2005. Antibacterial activity of *Cuminum cyminum* L. and *Carum carvi* L. essential oils. *Journal of agricultural and food chemistry* **53**, 57–61.

Jalali I, Parashar RD, 1995. Biocontrol of *Xanthomonas campestris* pv. *campestris* in *Brassica juncea* with phylloplane antagonist. *Plant Disease Research* **10**, 145–147.

Jehle AK, Lipschis M, Albert M *et al.*, 2013. The receptor-like protein ReMAX of *Arabidopsis* detects the microbe-associated molecular pattern eMax from Xanthomonas. *The Plant cell* **25**, 2330–2340.

Jensen BD, Vicente JG, Manandhar HK, Roberts SJ, 2010. Occurrence and Diversity of *Xanthomonas campestris* pv. *campestris* in Vegetable Brassica Fields in Nepal. *Plant Disease* **94**, 298–305.

Jeppesen MG, Navratil T, Spremulli LL, Nyborg J, 2005. Crystal structure of the bovine mitochondrial elongation factor Tu.Ts complex. *The Journal of biological chemistry* **280**, 5071–5081.

Jones JDG, Dangl JL, 2006. The plant immune system. *Nature* **444**, 323–329.

Kamoun S, Kamdar H V, Tola E, Kado CI, 1992. Incompatible Interactions Between Crucifers and *Xanthomonas campestris* Involve a Vascular Hypersensitive Response : Role of the *hrpX* Locus. *Molecular Plant-Microbe Interactions* **5**, 22–33.

Kim S, Cho Y-J, Song E-S *et al.*, 2016. Time-resolved pathogenic gene expression analysis of the plant pathogen *Xanthomonas oryzae* pv. *oryzae*. *BMC Genomics* **17**, 345.

Kocks CG, 1996. Cabbage Refuse Piles as Sources of Inoculum for Black Rot Epidemics. *Plant Disease* **80**, 789–792.

Kocks CG, Zadoks J-C, Ruissen TA, 1998. Response of black rot in cabbage to spatial distribution of inoculum. *European Journal of Plant Pathology* **104**, 713–723.

Kocks CG, Zadoks JC, Ruissen MA, 1999. Spatio-temporal development of black rot (*X. campestris* pv. *campestris*) in cabbage in relation to initial inoculum levels in field plots in The Netherlands. *Plant Pathology* **48**, 176–188.

Kunze G, 2004. The N Terminus of Bacterial Elongation Factor Tu Elicits Innate Immunity in *Arabidopsis* Plants. *The plant cell* **16**, 3496–3507.

Lannou C, 2012. Variation and selection of quantitative traits in plant pathogens. *Annual Review of Phytopathology* **50**, 319–338.

Lapouge K, Schubert M, Allain FH-T, Haas D, 2008. Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Molecular microbiology* **67**, 241–53.

Lee J, Izzah NK, Jayakodi M *et al.*, 2015. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. *BMC Plant Biology* **15**, 1–11.

Li R-F, Lu G-T, Li L *et al.*, 2013. Identification of a putative cognate sensor kinase for the two-component response regulator HrpG, a key regulator controlling the expression of the hrp genes in *Xanthomonas campestris* pv. *campestris*. *Environmental microbiology* **16**, 2053–2071.

Lindgren PB, Peet RC, Panopoulos NJ, 1986. Gene cluster of Pseudomonas syringae pv. “phaseolicola”

controls pathogenicity of bean plants and hypersensitivity of nonhost plants. *Journal of Bacteriology* **168**, 512–522.

Liu W, Yu Y-HH, Cao S-YY *et al.*, 2013. Transcriptome profiling of *Xanthomonas campestris* pv. *campestris* grown in minimal medium MMX and rich medium NYG. *Research in Microbiology* **164**, 466–479.

Lugo AJ, Elibox W, Jones JB, Ramsubhag A, 2013. Copper resistance in *Xanthomonas campestris* pv. *campestris* affecting crucifers in Trinidad. *European Journal of Plant Pathology* **136**, 61–70.

Massomo SMS, Mortensen CN, Mabagala RB, Newman M-A, Hockenhull J, 2004. Biological Control of Black Rot (*Xanthomonas campestris* pv. *campestris*) of Cabbage in Tanzania with Bacillus strains. *Journal of Phytopathology* **152**, 98–105.

Massomo SMS, Nielsen H, Mabagala RB, Mansfeld-Giese K, Hockenhull J, Mortensen CN, 2003. Identification and Characterisation of *Xanthomonas campestris* pv. *campestris* Strains from Tanzania by Pathogenicity Tests, Biolog, rep-PCR and Fatty Acid Methyl Ester Analysis. *European Journal of Plant Pathology* **109**, 775–789.

McCarthy Y, Ryan RP, O'Donovan K *et al.*, 2008. The role of PilZ domain proteins in the virulence of *Xanthomonas campestris* pv. *campestris*. *Molecular plant pathology* **9**, 819–824.

McCulloch L, 1929. A bacterial leaf spot of horse-radish caused by *Bacterium campestris* var. *armoraciae*, n. var. *Journal of Agricultural Research* **38**, 269–287.

McDonald C, Inohara N, Nuñez G, 2005. Peptidoglycan signaling in innate immunity and inflammatory disease. *The Journal of biological chemistry* **280**, 20177–20180.

Meier D, 1934. A cytological study of the early infection stages of the black rot of cabbage. *Bulletin of the Torrey Botanical Club* **61**, 173–190.

Mew TW, Natural MP, 1993. Management of *Xanthomonas* diseases. In: Swings JG,, Civerolo EL, eds. *Xanthomonas*. London: Chapman & Hall, 341–362.

Moller W, Schroth M, Thomson S, 1981. The Scenario of Fire Blight and Streptomycin Resistance. *Plant Disease* **65**, 563–568.

Monteiro L, Mariano RLR, Souto-Maior AM, 2005. Antagonism of *Bacillus* spp. Against *Xanthomonas campestris* pv. *campestris*. *Brazilian Archives of Biology and Technology* **48**, 23–29.

Newman M-A, Sundelin T, Nielsen JT, Erbs G, 2013. MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Frontiers in Plant Science* **4**, 1–14.

Nguyen HP, Chakravarthy S, Velásquez AC *et al.*, 2010. Methods to study PAMP-triggered immunity using tomato and *Nicotiana benthamiana*. *Molecular plant-microbe interactions : MPMI* **23**, 991–999.

O'Brien EJ, Bennett P, 1972. Structure of Straight Flagella from Mutant. *Journal of Molecular Biology* **70**, 133–152.

Pal KK, Gardener BMS, 2006. Biological Control of Plant Pathogens. *The Plant Health Instructor*, 1–25.

Pammel L, 1895. Bacteriosis of rutabaga (*Bacillus campestris* n sp.). *Iowa State College Agricultural Experiment Station Bulletin* **16**, 130–134.

Pandey SP, Somssich IE, 2009. The role of WRKY transcription factors in plant immunity. *Plant physiology* **150**, 1648–1655.

Panstruga R, Parker JE, Schulze-Lefert P, 2009. SnapShot: Plant Immune Response Pathways. *Cell* **136**, 6–8.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D, 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **57**, 2881–2887.

Parkinson N, Cowie C, Heeney J, Stead D, 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *International Journal of Systematic and Evolutionary Microbiology* **59**, 264–274.

Pichard B, Thouvenot D, 1999. Effect of Bacillus polymyxa seed treatments on control of black-rot and damping-off of cauliflower. *Seed Science and Technology* **27**, 455–465.

Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD, Cociancich S, 2015. What makes *Xanthomonas albilineans* unique amongst xanthomonads? *Frontiers in plant science* **6**, 1–7.

Pinto-Ganhão J, 1962. *Xanthomonas campestris* (Pammel) Dowson - Uma nova bacteriose para Portugal. *Agros* **45**, 167–168.

Qian W, Han Z-J, He C, 2008. Two-component signal transduction systems of *Xanthomonas* spp.: a lesson from genomics. *Molecular plant-microbe interactions : MPMI* **21**, 151–161.

Qian W, Jia Y, Ren S-X *et al.*, 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome research* **15**, 757–767.

Rademaker J, Hoste B, Louws FJ *et al.*, 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *International journal of*

systematic and evolutionary microbiology **50**, 665–677.

Rademaker JLW, Louws FJ, Schultz MH *et al.*, 2005. A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology* **95**, 1098–1111.

Reinhardt JA, Baltrus DA, Nishimura MT, Jeck WR, Jones CD, Dangl JL, 2009. *De novo* assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. *Genome research* **19**, 294–305.

Richardson JK, 1945. Black rot of rutabagas. *Scientific Agriculture* **25**, 415–425.

Roberts SJ, Brough J, Hunter PJ, 2007. Modelling the spread of *Xanthomonas campestris* pv. *campestris* in module-raised brassica transplants. *Plant Pathology* **56**, 391–401.

Roberts SJ, Koenraadt H, 2006. 7-019a: Detection of *Xanthomonas campestris* pv. *campestris* on Brassica spp. In: *International Rules for Seed Testing*. Bassersdorf: International Seed Testing Association (ISTA), 1–16.

Robeson DJ, Bretschneider KE, Gonella MP, 1989. A hydathode inoculation technique for the simulation of natural black rot infection of cabbage by Xanthomonas campestris pv. campestris. *Annals of Applied Biology* **115**, 455–459.

Romeo T, Vakulskas CA, Babitzke P, 2013. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environmental microbiology* **15**, 313–324.

Roux B, Bolot S, Guy E *et al.*, 2015. Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. *BMC genomics* **16**, 1–19.

Ryan RP, Fouhy Y, Lucey JF *et al.*, 2007. Cyclic di-GMP signalling in the virulence and environmental adaptation of *Xanthomonas campestris*. *Molecular Microbiology* **63**, 429–442.

Ryan RP, Vorhölter F-J, Potnis N *et al.*, 2011. Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nature reviews. Microbiology* **9**, 344–355.

Satish S, Raveesha K a., Janardhana GR, 1999. Antibacterial activity of plant extracts on phytopathogenic *Xanthomonas campestris* pathovars. *Letters in Applied Microbiology* **28**, 145–147.

Sauer DB, 1986. Disinfection of Seed Surfaces with Sodium Hypochlorite. *Phytopathology* **76**, 745–749.

Schaad NW, 1974. Survival of *Xanthomonas campestris* in Soil. *Phytopathology* **64**, 1518–1520.

Schaad NW, Alvarez A, 1993. *Xanthomonas campestris* pv. *campestris*: cause of black rot of crucifers.

In: Swings JG,, Civerolo EL, eds. *Xanthomonas*. London: Chapman & Hall, 51–55.

Schaad NW, Dianese JC, 1981. Cruciferous weeds as sources of inoculum of *Xanthomonas campestris* in black rot crucifers. *Phytopathology* **71**, 1215–1220.

Schaad NW, Gabrielson RL, Mulanax MW, 1980. Hot acidified cupric acetate soaks for eradication of *Xanthomonas campestris* from crucifer seeds. *Applied and Environmental Microbiology* **39**, 803–807.

Schaad NW, Jones JB, Chun W, 2001. *Laboratory Guide for Identification of Plant Pathogenic Bacteria*. St. Paul: APS Press.

Schaad NW, Vidaver AK, Lacy GH, Rudolph K, Jones JB, 2000. Evaluation of proposed amended names of several pseudomonads and xanthomonads and recommendations. *Phytopathology* **90**, 208–213.

Schneider DJ, Collmer A, 2010. Studying plant-pathogen interactions in the genomics era: beyond molecular Koch's postulates to systems biology. *Annual review of phytopathology* **48**, 457–479.

Schultz T, 1986. Control of *Xanthomonas campestris* pv. *campestris* in crucifer seed with slurry treatments of calcium hypochlorite. *Plant Disease* **70**, 1027–1030.

Sequeira L, 1983. Mechanisms of induced resistance in plants. *Annual review of microbiology* **37**, 51–79.

Sidhu VK, Vorhölter F-J, Niehaus K, Watt SA, 2008. Analysis of outer membrane vesicle associated proteins isolated from the plant pathogenic bacterium *Xanthomonas campestris* pv. *campestris*. *BMC microbiology* **8**, 1–16.

Silipo A, Molinaro A, Sturiale L *et al.*, 2005. The elicitation of plant innate immunity by lipooligosaccharide of *Xanthomonas campestris*. *The Journal of biological chemistry* **280**, 33660–33668.

da Silva ACR, Ferro JA, Reinach FC *et al.*, 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463.

Simões THN, Gonçalves ER, Rosato YB, Mehta A, 2007. Differentiation of *Xanthomonas* species by PCR-RFLP of rpfB and atpD genes. *FEMS microbiology letters* **271**, 33–39.

Simpson RB, Johnson LJ, 1990. *Arabidopsis thaliana* as a host for *Xanthomonas campestris* pv. *campestris*. *Molecular Plant-Microbe Interactions* **3**, 233–237.

Simpson AJG, Reinach FC, Arruda P *et al.*, 2000. The genome sequence of the plant pathogen *Xylella*

fastidiosa. *Nature* **406**, 151–157.

Singh D, Raghavendra BT, Rathaur Singh P, Singh H, Raghuwanshi R, Singh RP, 2014. Detection of black rot disease causing pathogen *Xanthomonas campestris* pv. *campestris* by bio-PCR from seeds and plant parts of cole crops. *Seed Science and Technology* **42**, 36–46.

Sinha RK, 2004. *Modern Plant Physiology*. Pangbourne: Alpha Science International.

Smith EF, 1898. The black rot of the cabbage. *U.S. Department of Agriculture, Farmers' Bulletin* **68**, 1–21.

Sokmen A, Gulluce M, Akpulat HA *et al.*, 2004. The in vitro antimicrobial and antioxidant activities of the essential oils and methanol extracts of endemic *Thymus spathulifolius*. *Food Control* **15**, 627–634.

Staub T, Williams PH, 1972. Factors influencing black rot lesion development in resistant and susceptible cabbage. *Phytopathology* **62**, 722–728.

Stockwell VO, Duffy B, 2012. Use of antibiotics in plant agriculture. *Revue scientifique et technique (International Office of Epizootics)* **31**, 199–210.

Studholme DJ, Glover RH, Boonham N, 2011. Application of high-throughput DNA sequencing in phytopathology. *Annual review of phytopathology* **49**, 87–105.

Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF, 2006. Within-species flagellin polymorphism in *Xanthomonas campestris* pv. *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2-dependent defenses. *The Plant cell* **18**, 764–779.

Tang D-J, Li X-J, He Y-Q, Feng J-X, Chen B, Tang J-L, 2005. The zinc uptake regulator Zur is essential for the full virulence of *Xanthomonas campestris* pv. *campestris*. *Molecular plant-microbe interactions : MPMI* **18**, 652–658.

Tang X, Xiao Y, Zhou J, 2006. Regulation of the Type III Secretion System in Phytopathogenic Bacteria. *Molecular Plant-Microbe Interactions* **19**, 1159–1166.

Tao J, Li C, Luo C, He C, 2014. RavA/RavR two-component system regulates *Xanthomonas campestris* pathogenesis and c-di-GMP turnover. *FEMS microbiology letters* **358**, 81–90.

Taylor JD, Conway J, Roberts SJ, Astley D, Vicente JG, 2002. Sources and Origin of Resistance to *Xanthomonas campestris* pv. *campestris* in Brassica Genomes. *Phytopathology* **92**, 105–111.

Taylor AG, Salanenka YA, 2012. Seed treatments: phytotoxicity amelioration and tracer uptake. *Seed Science Research* **22**, S86–S90.

Tjamos EC, Tjamos SE, Antoniou PP, 2010. Biological Management of Plant Diseases : Highlights on Research and Application of Biological Management. *Journal of Plant Pathology* **92**, 17–22.

Tsuda K, Katagiri F, 2010. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology* **13**, 459–465.

U N, 1935. Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7**, 389–452.

Vauterin L, Hoste B, Kersters K, Swings J, 1995. Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology* **45**, 472–489.

Vauterin L, Hoste B, Yang P, Alvarez A, Kersters K, Swings J, 1993. Taxonomy of the genus *Xanthomonas*. In: Swings J,, Civerolo EL, eds. *Xanthomonas*. London: Chapman & Hall, 157–192.

Vauterin L, Rademaker J, Swings J, 2000. Synopsis on the taxonomy of the genus Xanthomonas. *Phytopathology* **90**, 677–682.

Vauterin L, Swings J, Kersters K *et al.*, 1990. Towards an Improved Taxonomy of *Xanthomonas*. *International Journal of Systematic Bacteriology* **40**, 312–316.

Vicente JG, 2000. Diversidade da *Xanthomonas campestris* pv. *campestris* (Pammel) Dowson e caracterização da resistência genética em Brassica spp.. Universidade Técnica de Lisboa.

Vicente JG, 2004. *A Podridão Negra das Crucíferas* (G Lopes, Ed,). Alcobaça: COTHN Centro Operativo Tecnológico Hortofruticola.

Vicente JG, Conway J, Roberts SJ, Taylor JD, 2001. Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathovars. *Phytopathology* **91**, 492–499.

Vicente JG, Everett B, Roberts SJ, 2006. Identification of isolates that cause a leaf spot disease of Brassicas as *Xanthomonas campestris* pv. *raphani* and pathogenic and genetic comparison with related pathovars. *Phytopathology* **96**, 735–745.

Vicente JG, Holub EB, 2013. *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular plant pathology* **14**, 2–18.

Vicente JG, Taylor JD, Sharpe a G, Parkin I a P, Lydiate DJ, King GJ, 2002. Inheritance of Race-Specific Resistance to *Xanthomonas campestris* pv. *campestris* in Brassica Genomes. *Phytopathology* **92**, 1134–1341.

Voloudakis AE, Reignier TM, Cooksey DA, 2005. Regulation of resistance to copper in Xanthomonas axonopodis pv. vesicatoria. *Applied and environmental microbiology* **71**, 782–789.

Vorhölter F-J, Schneiker S, Goesmann A *et al.*, 2008. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *Journal of biotechnology* **134**, 33–45.

Wallis FM, Rijkenberg FHJ, Joubert JJ, Martin MM, 1973. Ultrastructural histopathology of cabbage leaves infected with *Xanthomonas campestris*. *Physiological Plant Pathology* **3**, 371–378.

Wang L-H, He Y, Gao Y *et al.*, 2003. A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Molecular Microbiology* **51**, 903–912.

Wang G, Roux B, Feng F *et al.*, 2015. The Decoy Substrate of a Pathogen Effector and a Pseudokinase Specify Pathogen-Induced Modified-Self Recognition and Immunity in Plants. *Cell Host and Microbe* **18**, 285–295.

Wengelnik K, Van Den Ackerveken G, Bonas U, 1996. HrpG, a key hrp regulatory protein of *Xanthomonas campestris* pv. *vesicatoria* is homologous to two-component response regulators. *Molecular Plant-Microbe Interactions* **9**, 704–712.

Westermann AJ, Gorski S a, Vogel J, 2012. Dual RNA-seq of pathogen and host. *Nature reviews. Microbiology* **10**, 618–630.

Williams PH, 1980. Black Rot: A Continuing Threat to World Crucifers. *Plant Disease* **64**, 736–742.

Van Der Wolf JM, Van Der Zouwen PS, 2010. Colonization of Cauliflower Blossom (*Brassica oleracea*) by *Xanthomonas campestris* pv. *campestris*, via Flies (*Calliphora vomitoria*) Can Result in Seed Infestation. *Journal of Phytopathology* **158**, 726–732.

Wulff EG, Mguni CM, Mortensen CN, Keswani CL, Hockenhull J, 2002. Biological Control of Black Rot (*Xanthomonas campestris* pv. *campestris*) of Brassicas with an Antagonistic Strain of *Bacillus subtilis* in Zimbabwe. *European Journal of Plant Pathology* **108**, 317–325.

Yadeta KA, Thomma BPH, 2013. The xylem as battleground for plant hosts and vascular wilt pathogens. *Frontiers in plant science* **4**, 1–12.

Young JM, Dye DW, Bradbury JF, Panagopoulos CG, Robbs CF, 1978. A proposed nomenclature and classification for plant pathogenic bacteria. *New Zealand Journal of Agricultural Research* **21**, 153–177.

Young JM, Park D-C, Shearman HM, Fargier E, 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Systematic and Applied Microbiology* **31**, 366–377.

Zaccardelli M, Campanile F, Spasiano A, Merighi M, 2007. Detection and identification of the crucifer

pathogen, *Xanthomonas campestris* pv. *campestris*, by PCR amplification of the conserved Hrp/type III secretion system gene *hrcC*. *European Journal of Plant Pathology* **118**, 299–306.

Zhang S-S, He Y-Q, Xu L-M *et al.*, 2008. A putative colR(XC1049)-colS(XC1050) two-component signal transduction system in *Xanthomonas campestris* positively regulates hrpC and hrpE operons and is involved in virulence, the hypersensitive response and tolerance to various stresses. *Research in microbiology* **159**, 569–578.

Zhang J, Yin Z, White F, 2015. TAL effectors and the executor R genes. *Frontiers in plant science* **6**, 1–9.

Zipfel C, 2014. Plant pattern-recognition receptors. *Trends in Immunology* **35**, 345–351.

Zipfel C, Kunze G, Chinchilla D *et al.*, 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell* **125**, 749–760.

II. Characterization of *Xanthomonas campestris* isolates collected in Portugal

THIS PART IS DIVIDED INTO THE FOLLOWING CHAPTERS:

THE RESULTS DESCRIBED IN THIS PART HAVE BEEN INCLUDED IN THE FOLLOWING PUBLICATIONS:

- **Cruz, J.**, Tenreiro, R., Cruz, L. (2015). First Report of *Xanthomonas campestris* pv. *raphani* Causing Leaf Spot Disease of *Brassica oleracea* in Portugal. Plant Disease 99 (2): 282.
- **Cruz, J.**, Tenreiro, R., Cruz, L. (2017) Assessment of Diversity of *Xanthomonas campestris* Pathovars Affecting Cruciferous Plants in Portugal And Disclosure of Two Novel *X. campestris* pv. *campestris* Races. Journal of Plant Pathology 99 (2): 403-414.
- **Cruz, J.,** Tenreiro, R., Cruz, L. (2017) Inference of phylogenetic diversity and population structure of Xanthomonas campestris affecting Brassicaceae using a MLST-based approach. Plant Pathology, doi:10.1111/ppa.12791

1 Diversity of *Xanthomonas campestris* pathovars and *X. campestris* pv. *campestris* races affecting cruciferous plants in Portugal

1.1 Introduction

Xanthomonas campestris is currently divided according to host specificity into three pathovars: *X. campestris* pv. *campestris* (*Xcc*), *X. campestris* pv. *raphani* (*Xcr*) and *X. campestris* pv. *incanae* (*Xci*), causing three distinct diseases (Fargier & Manceau, 2007).

Black rot disease, caused by *Xcc*, is the most important Brassicaceae disease worldwide, affecting economically important crops as wells as ornamentals and weeds. This vascular disease, which ultimately leads to plant death, causes typical V-shaped lesions on the leaf margins, following darkening of the veins. *Xcr* causes leaf spot disease on both Brassicaceae and Solanaceae hosts and is characterized by small dark spots that develop to brown lesions often surrounded by chlorotic halos. Bacterial blight disease, caused by *Xci*, has been described in ornamental cruciferous plants exhibiting yellowing and wilting of leaves.

As for several other phytopathogenic bacteria, a race structure has been established for *Xcc* and *Xcr*, based on the pathogenic profiles for several hosts. Race differentiation among *Xcc* was first proposed in the early 1990's (Kamoun *et al.*, 1992) and since then it has become increasingly complex.

Currently *Xcc* is divided into 9 races, of which races 1 and 4 are considered prevalent worldwide (Fargier & Manceau, 2007; Vicente & Holub, 2013). Race typing of *Xcr* has allowed the establishment of 3 races, with races 1 and 3 prevailing (Vicente *et al.*, 2006). Although no race structure has been adopted for *Xci*, several studies have established the existence of host specificity within this pathovar (Vicente *et al.*, 2001; Fargier & Manceau, 2007).

Gene-for-gene models, based on the interaction of matching resistance-virulence genes, have been proposed to explain the existence of races within *Xcc* and *Xcr* (Vicente *et al.*, 2002, 2006; He *et al.*, 2007). However, due to the lack of molecular data from both host and pathogen, validation of such models is still pending. Despite of its convenience for diagnosis and epidemiological studies, the use of race structure within *X. campestris* pathovars is still hindered, mostly due to the lack of molecular evidences justifying existing relationships between *Xcc* strains and differential hosts.

Sequencing of constitutive genes have been successful in distinguishing isolates at pathovar level (Vicente *et al.*, 2006; Parkinson *et al.*, 2007, 2009). Molecular research of *Xcc* gained thrust with the release of fully sequenced genomes of *Xcc* (da Silva *et al.*, 2002; Qian *et al.*, 2005; Vorhölter *et al.*,

2008; Bolot *et al.*, 2013a,b) and *Xcr* isolates (Bogdanove *et al.*, 2011). These reference genomes have revealed several virulence genes required for plant association/pathogenicity, based on homology with known genes from other *Xanthomonas* species.

In Portugal, where brassica crops are socio-economically very important, *X. campestris* causing black rot in a wide range of hosts, such as cabbage, mustard, turnip and radish, was first recorded in 1961 (Pinto-Ganhão, 1962). In the present study, with the aim of bringing new insights on the characterization of *X. campestris* pathovars occurring in Portugal, a set of 33 isolates collected in Portugal and previously identified as *X. campestris* was studied, combining phenotypic and genomic data. In addition to biological assays for pathovar and race determination (for *Xcc* only), phylogenetic analysis based on constitutive *gyrB* gene and PCR screening of selected known virulence genes were carried out.

1.2 Materials and Methods

1.2.1 Bacterial isolates

A set of 33 bacterial isolates from the Portuguese collection of phytopathogenic bacteria (CPBF, Coleção Portuguesa de Bactérias Fitopatogénicas) were used in this study (Table 5). Before use, isolates were grown for 48h at 27°C in Yeast Dextrose Chalk Agar media (Lelliot and Stead, 1987). For DNA extraction, to be used in PCR screening of virulence genes, as well as for sequencing of gyrB gene, a water suspension from each bacterial isolate (OD_{600}=0.1) was heated during 7 min at 96°C and then stored at -20°C, for future usage.

1.2.2 Biological tests for pathovar and race determination

In order to assess the pathogenicity of the isolates and to establish *X. campestris* pathovars, initial inoculations of Savoy cabbage (*B. oleracea* var. *sabauda* cv. 'Wirosa F1' – Sommers Seeds, BE), Portuguese cabbage (*B. oleracea var. costata* cv. 'Beira F1' – Bejo Zaden, NL) and Galega kale (*B. oleracea convar. acephala var. sabellica* cv. 'Bonanza F1' – Bejo Zaden, NL) were performed, by spraying a bacterial suspension (OD_{600}=0.1) directly onto the leaf surface. Atypical isolates, i.e. causing leaf spots or weakly pathogenic, were further inoculated on *Solanum lycopersicum*, *Raphanus sativus*, *Mathiola incana*, *Erysimum cheiri* (syn. *Cheiranthus cheiri*) and *E. linifolium*.

After inoculation, plantlets were maintained in closed plastic bags for 48h in order to promote the infection process. Plants were kept for 15 days at 24/18°C with a 16h-light/8h-dark photoperiod (RH>80%) and checked daily for symptom development.

For race-typing of *Xcc* isolates, biological assays were carried out on *Brassica* spp. cultivars (Vicente *et al.*, 2001; Fargier & Manceau, 2007). Inoculations were performed in plantlets with at least

three true leaves, in a total of 20 inoculation points per host by clipping the leaf margins with mouse-tooth forceps carrying bacterial suspension (OD_{600}=0.1). After inoculation, plants were kept as described above. Typical symptoms were checked each 2-3 days, for 2 weeks. For all biological assays, the type strain for *Xcc* (CFBP5241^T) was used as a positive control and water as a negative control.

1.2.3 Phylogenetic analysis

Phylogenetic characterization of Portuguese isolates was achieved by *gyrB* gene sequencing. Amplification was performed using the primer set X.gyr.fsp.s1/X.gyr.rsp3 and following amplification conditions previously described (Parkinson *et al.*, 2007). The size of the amplicons was confirmed by 1.5% agarose gel electrophoresis in 1X TAE at 3V.cm^{-1} during 1h and PCR products were then purified using JetQuick PCR Purification Spin Kit (GENOMED). Sequencing reactions were performed using an automatic sequencer CEQ 2000-XL (Beckman Coulter). Sequences were analysed and corrected using Chromas Lite 2.01 software (www.technelysium.com.au) and compared by nucleotide BLAST with those available at GenBank (NCBI). Sequences were edited using BioEdit 7.0.9.0 software (www.mbio.ncsu.edu/BioEdit/bioedit.html) and their complete alignment was performed using MEGA 5.1 software (Tamura *et al.*, 2011). This software was also used to construct a phylogenetic tree, using Neighbor-Joining method. In addition to the sequences obtained in this study, the sequences of other xanthomonads were included in the phylogeny. Bootstrap values were calculated from 1000 replications.

Table 5 - Portuguese *Xanthomonas campestris* isolates used in this study.

CPBF number	Isolation [a]			Pathovar determination [b]							*Xcc* race determination [c]						PVP[d]
	Host	Year	County	1	2	3	4	5	6	7	8	9	10	11	12	13	
Xanthomonas campestris pv. *campestris*																	
RACE 4																	
46	*B. oleracea convar. acephala* var. *sabellica*	2003	Sintra	v+	v+	v+	nd	nd	nd	nd	v+	-	-	-	v+	v+	A
RACE 6																	
208	*B. oleracea* var. *costata*	2003	Sintra	v+	-	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
213	*B. oleracea* var. *costata*	2004	Sintra	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
278	*B. oleracea*	1992		v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
279	*B. oleracea*			v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
589	*B. oleracea* convar. *botrytis* var. *botrytis*	2005	Santarém	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
667	*B. oleracea*			v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
1126	*B. oleracea*	2002	Pombal	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
1135	*B. oleracea* convar. *capitata* var. *alba*	2002	Odemira	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
1136	*B. oleracea* convar. *capitata* var. *alba*	2002	Odemira	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
1175	*B. oleracea* convar. *botrytis* var. *botrytis*	2006	Peniche	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
1176	*B. oleracea* convar. *botrytis* var. *botrytis*	2006	Peniche	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	v+	v+	v+	A
RACE 7																	
140	*B. oleracea* var. *costata*	2003	Almada	v+	v+	-	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
210	*B. oleracea* convar. *botrytis* var. *botrytis*	2004	Coruche	v+	v+	-	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
211	*B. oleracea* convar. *botrytis* var. *botrytis*	2004	Coruche	v+	-	-	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
212	*B. oleracea* var. *costata*	2004	Sintra	v+	-	-	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
216	*B. oleracea* convar. *botrytis* var. *botrytis*	2005	Santarém	v+	v+	-	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
329	*B. oleracea* var. *costata*	2004	Sintra	v+	v+	-	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
332	*B. oleracea* var. *costata*	2004	Sintra	v+	v+	v+	nd	nd	nd	nd	v+	v+	v+	-	v+	v+	A
668	*B. oleracea*	1998		-	-	-	(ls+)	v+	-	v+	v+	v+	v+	-	v+	v+	A
824	*B. oleracea*			v+	v+	v+	nd	nd	nd	nd	v+	v+	+	-	v+	v+	A
RACE 10																	
330	*B. oleracea* var. *costata*	2004	Sintra	v+	v+	v+	nd	nd	nd	nd	v+	v+	-	-	v+	v+	A
331	*B. oleracea* var. *costata*	2004	Sintra	v+	v+	-	nd	nd	nd	nd	v+	v+	-	-	v+	v+	A

CPBF number	Isolation [a]			Pathovar determination [b]							*Xcc* race determination [c]						PVP[d]
	Host	Year	County	1	2	3	4	5	6	7	8	9	10	11	12	13	
604	*B. oleracea* convar. *capitata* var. *sabauda*	2005	Santarém	v+	v+	v+	nd	nd	nd	nd	v+	v+	-	-	v+	v+	A
RACE 11																	
489	*B. oleracea* var. *costata*	1997	Loures	-	-	-	-	v+	-	v+	v+	v+	-	v+	v+	v+	A
1198	*B. kaber* var. *pinnatifida*	2007	Lisboa	-	-	v+	nd	nd	nd	nd	v+	v+	-	v+	v+	v+	A
Xanthomonas campestris pv. *raphani*																	
143	*B. oleracea* convar. *capitata* var. *alba*	2003	Sintra	ls +	-	-	ls+	-	-	ls+	nd	nd	nd	nd	nd	nd	B
207	*B. oleracea* convar. *botrytis* var. *botrytis*	2003	Sintra	ls+	ls+	ls+	ls+	-	-	ls+	nd	nd	nd	nd	nd	nd	A
209	*B. oleracea* var. *costata*	2004	Sintra	ls+	ls+	ls+	ls+	-	-	ls+	nd	nd	nd	nd	nd	nd	A
1171	*B. oleracea* convar. *capitata* var. *sabauda*	2006	Santarém	ls+	ls+	ls+	ls+	-	-	ls+	nd	nd	nd	nd	nd	nd	A
Xanthomonas campestris pv. *incanae*																	
1257	*Erysimum* sp.	2008	Montijo	-	-	(v+)	(ls+)	-	b+	-	nd	nd	nd	nd	nd	nd	A
Non-pathogenic *Xanthomonas campestris*																	
394	*Ranunculus* sp.	2005	Montijo	-	-	-	-	-	-	-	nd	nd	nd	nd	nd	nd	C
Xanthomonas sp.																	
47	*Rorippa nasturtium-aquaticum*	2003	Sintra	-	-	-	ls+	-	-	ls+	nd	nd	nd	nd	nd	nd	D

[a] - Year of isolation and county of origin given when known

[b] - 1 - *B. oleracea* cv. 'Wirosa'; 2 - *B. oleracea* cv. 'Beira'; 3 - *B. oleracea* cv. 'Bonanza'; 4 - *S. lycopersicum; 5 - M. incana*; 6 – *Erysimum* spp.*; 7 - R. sativus;*

[c] - *8 - B. oleracea* cv. 'Wirosa'*; 9 - B. rapa* cv. 'Just Right Turnip'*; 10 - B. rapa* cv. 'Seven Top Turnip'*; 11 - B. juncea* cv. 'Florida Broad Leaf Mustard'*; 12 - B. carinata* PI 199947*; 13 - B. oleracea* cv. 'Miracle F1'

v, vascular infection; ls, leaf spots; b, blight symptoms; +, compatible interaction; -, incompatible interaction; ( ), weakly pathogenic; nd, not determined.

[d] - PVP (Partial Virulence Profile), based on the presence or absence of virulence genes, as described in Table 7.

1.2.4 PCR screening of known virulence genes

In order to partially characterize the virulence gene arsenal of each bacterial isolate, PCR screening of genes previously described as essential to virulence was conducted for the Portuguese isolates as well as for reference strain CFBP5241^T. The eight selected genes – *adkXcc* (Lu *et al.*, 2009), *dsbB* (Jiang *et al.*, 2008), *hpaR* (Wei *et al.*, 2007), *hrpF* (Berg *et al.*, 2005), *katG* (Jittawuttipoka *et al.*, 2009), *mip*-like (Zang *et al.*, 2007), *ppsA* (Tang *et al.*, 2005a) and *zur* (Tang *et al.*, 2005b) – were amplified using the primers described in Table 6, and the PCR reaction settings were defined according to the corresponding published procedures. All PCR reagents were obtained from Invitrogen. PCR products were analyzed by 1.5% agarose gel electrophoresis in 1X TAE at 3V.cm^{-1} for 90 min. The gels were stained and photographed as described for PCR fingerprinting. All PCR amplifications were repeated for result confirmation.

Table 6 - Primers used to assess the partial virulence profiles of Portuguese *Xanthomonas campestris* isolates.

Target Gene		Primers		Reference
Name	**Description**	**Name**	**Sequence (5' – 3')**	
adkXcc	adenosine kinase	XC_0690F XC_0690R	CGCACTGATCTGTGGTTCCCTCG TGGATCATGCCTCGCGCCC	Lu *et al.*, 2009
dsbB	disulfide bond formation protein B	XC3314-F XC3314-R	GGGGAATTCGTTGGGGTTTTCGGGCGCAATTC GGGGGATCCCCATACGCCTTGCGCCCACCGG	Jiang *et al.*, 2008
hpaR	hrp [hypersensitive response and pathogenicity]-associated regulator	2827MF 2827MR	CCCGGATCCGATCCAACCGCTAAACGCGTCG CCCAAGCTTTCGTCCGGGCTGCTGGCGCGCGC	Wei *et al.*, 2007
hrpF	hypersensitive response and pathogenicity protein F	DLH120 DLH125	CCGTAGCACTTAGTGCAATG GCATTTCCATCGGTCACGATTG	Berg *et al.* 2005
katG	catalase-peroxidase	BT2239 BT2240	TCTGCTTGCCACCGGACT TGTGGCAGGACCCGATCC	Jittawuttipoka *et al.*, 2009
mip-like	macrophage infectivity potentiator	2699PF 2699PR	GTCGGATCCATGAAGTTGCGTTCGATCGCGGTCGCTGTG CTCAAGCTTTTACTTGACGCTGACCAGCTTGATCTCGA	Zang *et al.*, 2007
ppsA	phosphoenolpyruvate synthase	ppsAF ppsAR	TATCCTGTGGTTGCATGAGCTACGC GGAACACATCTTCGTGCTTGAAGCC	Tang *et al.*, 2005b
zur	zinc uptake regulator	1430MF 1430MR	CAGGGATCCATGCCAACAGCTTCGTACGTGCGGT GGGAAGCTTGTCTGTGCCTGCGGTTGGAAGCCCA	Tang *et al.*, 2005a

1.3 Results

1.3.1 Phenotypic identification at pathovar level and race typing

Cultural and biochemical characterization, based on Lelliot and Stead (1987) and Vauterin *et al.* (1995), had previously allowed the putative identification of the set of 33 isolates as *X. campestris* (Cruz, 2009). For all isolates, typical yellow mucoid and convex colonies were recorded when isolates were grown for 48h on YDC (Yeast Dextrose Chalk) medium. All isolates could hydrolyse starch and esculin, as well as to produce levan. Isolates were oxidase negative, and unable to hydrolyse arginine and to use rhamnose, indol, inositol and sorbitol. This set of isolates were also able to use celobiose, trehalose, mannose, glucose, raffinose and sucrose.

Categorization of isolates at pathovar and race level using differential host cultivars and species is summarized in Table 5.

The inoculation of Savoy cabbage, Portuguese cabbage and Galega kale confirmed the presence of 24 *Xcc* isolates, causing typical black rot symptoms (Figure 9A, 9B). Four isolates caused leaf spots on *B. oleracea* cultivars, as well as on *S. lycopersicum* and *R. sativus*, confirming their identification as *Xcr* (Figure 9C, 9D). The inoculations on *B. oleracea* cultivars also revealed one weakly pathogenic isolate (CPBF1257) and 4 non-pathogenic ones (CPBF47, CPBF394, CPBF489 and CPBF668). Further inoculation of these five atypical isolates on *S. lycopersicum*, *R. sativus*, *E. cheiri* and *M. incana* showed that CPBF489 and CPBF668 produced typical black rot symptoms on *R. sativus* and *M. incana*, with CPBF668 causing some small leaf spots on *S. lycopersicum*. CPBF1257 caused typical bacterial blight symptoms on *Erysimum* spp. (Figure 9E, 9F), as well as a few small spots on *S. lycopersicum* leaves. CPBF47 also produced a small number of spots on *S. lycopersicum* and *R. sativus* leaves. CPBF394 was not pathogenic on any host tested.

Race-typing biological assays conducted for the 26 isolates causing black rot symptoms on at least one Brassicaceae host, enabled the allocation of 21 isolates to three previously described races – one isolate to race 4, 11 to race 6 and 9 to race 7. For the remaining five isolates, two new races were established, since they displayed pathogenicity profiles different from those of the nine currently described races (Fargier & Manceau, 2007). New race 10, that includes isolates CPBF330, CPBF331 and CPBF604, differs from race 4 by the ability to further successfully infect *B. rapa* cv. 'Just Right Turnip'. On the other hand, new race 11, represented by isolates CPBF489 and CPBF1198, differs from race 6 only by the absence of pathogenicity towards *B. rapa* cv. 'Seven Top Turnip'.

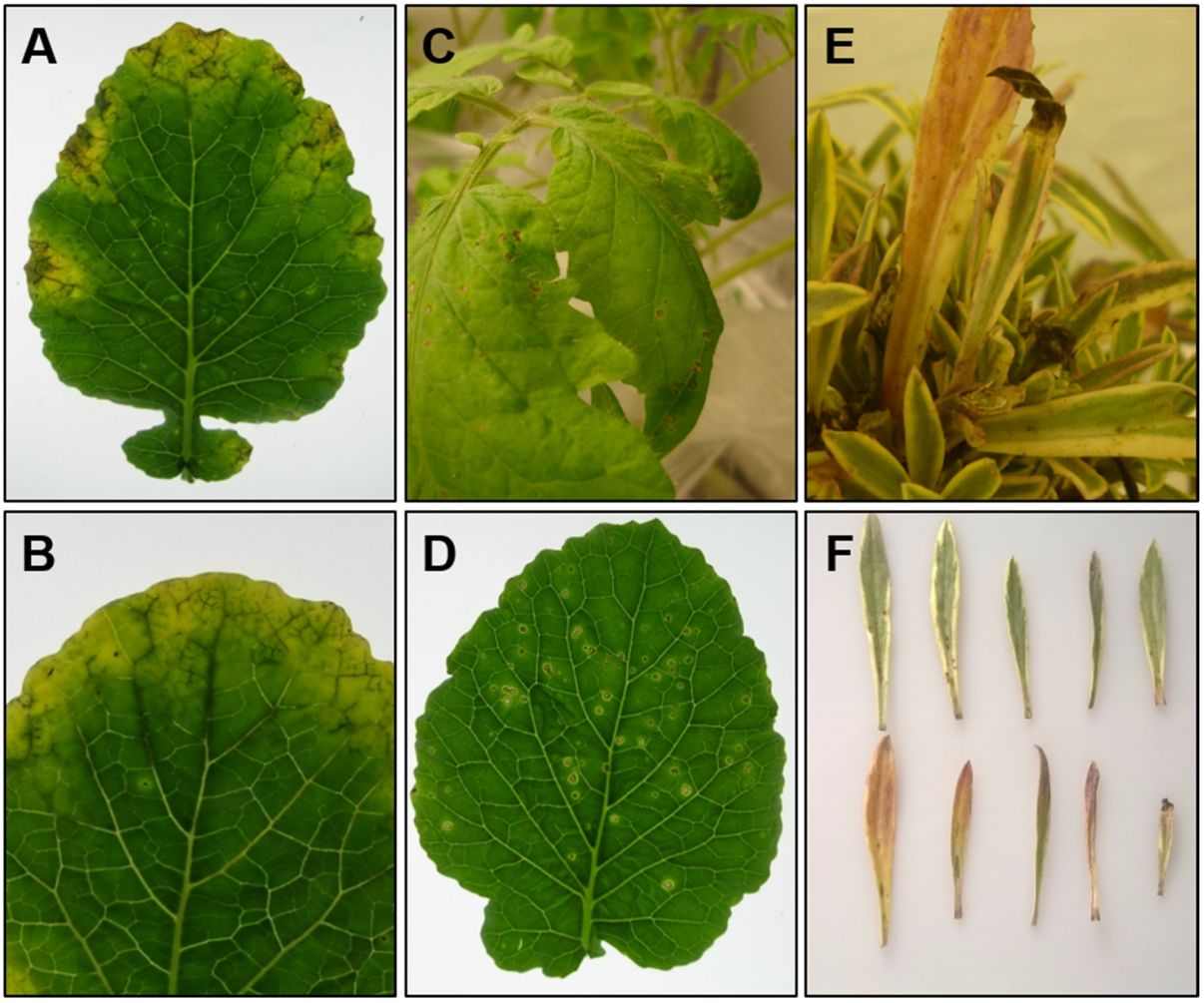


Figure 9 - Symptoms caused by Portuguese *Xanthomonas campestris* isolates. (A, B) V-shaped lesions with black veins on *Brassica oleracea* var. *sabauda* cv. 'Wirosa', caused by *X. campestris* pv. *campestris*. (C, D) Leaf spots on *Solanum lycopersicum* and *B. oleracea* var. *sabauda* cv. 'Wirosa', caused by *X. campestris* pv. *raphani*. (E, F) Blight on *Erysimum linifolium*, caused by *X. campestris* pv. *incanae*.

It is also important to notice that, for three *Xcc* isolates, the leaf spray inoculation on Savoy cabbage (cv. 'Wirosa') did not result in the production of typical black rot symptoms. Nevertheless, the leaf clipping inoculation used for race determination confirmed the vascular infection ability of these *Xcc* isolates. Moreover, a simple visual comparison between spray inoculations and previously performed clipping inoculations, revealed differences in symptom development, as highlighted in Figure 10. While spraying inoculation allowed a virulence-based discrimination of isolates, clipping inoculation produced very evident symptoms, similarly for all isolates, clearly masking their natural virulence. The effect of inoculation method on symptom development appears to be more significant for lower virulence isolates.

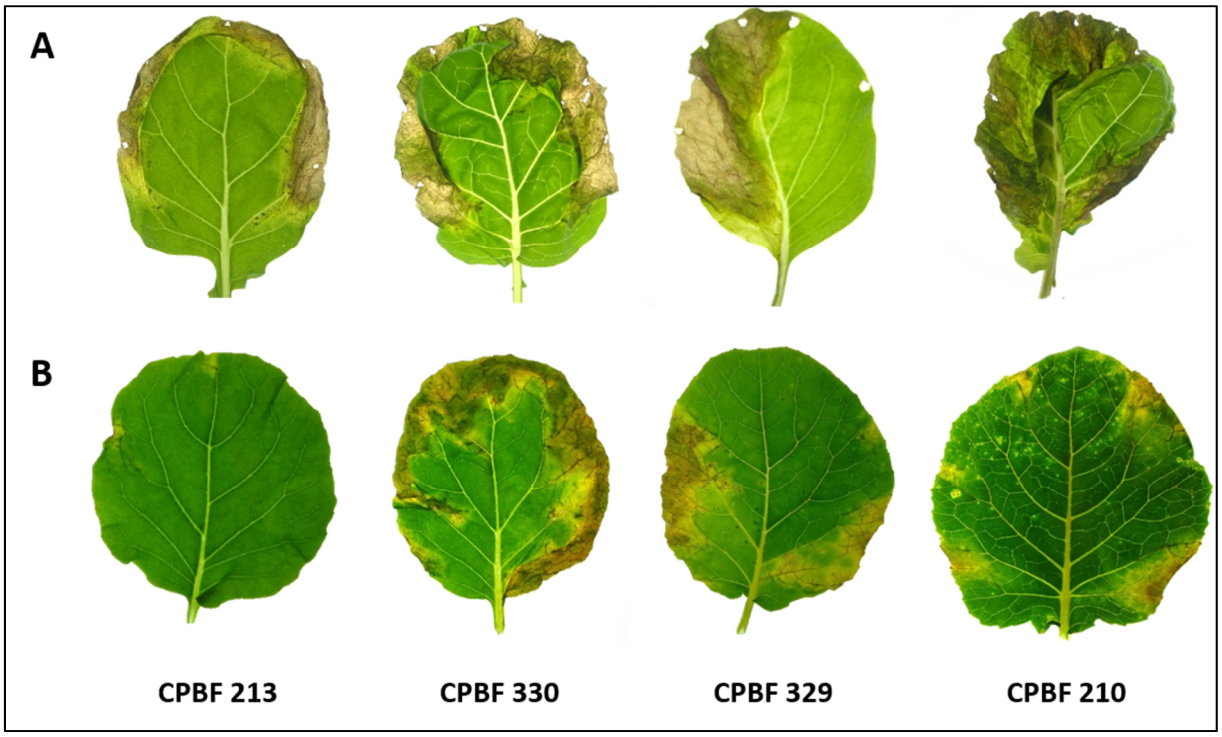


Figure 10 - Differences in symptoms 15 days after inoculation of different *Xanthomonas campestris* pv. *campestris* strains, using clipping (A) and spraying (B) inoculation methods.

1.3.2 Phylogenetic-based identification

The neighbour-joining phylogenetic tree based on the sequences of *gyrB* gene is shown in Figure 11. To assist in the phylogenetic allocation of the Portuguese isolates, sequences from different *Xanthomonas* species were also included in the phylogeny. Except for isolate CPBF47, all Portuguese isolates clustered together in a major clade with the type strains for the pathovars of *X. campestris* and separately of the remaining *Xanthomonas* species. Based on the topology of the tree it is possible to distinguish three separate clusters inside the *X. campestris* clade, corresponding to the three pathovars of this species. Cluster *Xcc* includes the type strain for *Xcc* and 24 isolates phenotypically allocated to this pathovar. Cluster *Xcr* includes the *Xcr* pathotype strain as well as three of the four isolates phenotypically identified as *Xcr*. Cluster *Xci* comprises the pathotype strain for *Xci* as well as isolate CPBF 1257, phenotypically identified as *X. campestris* pv. *incanae*.

Among the remaining isolates, the two *Xcc* pathogenic ones (CPBF208 and 213) are associated within the *Xci* cluster, whereas the non-pathogenic isolate CPBF394, together with *Xcr* isolate CPBF209, positioned externally to clusters *Xcr*, *Xcc* and *Xci*.

In terms of species level identification, all isolates except CPBF47 were confirmed by phylogenetic positioning to be accurately identified as *X. campestris*.

For isolate CPBF47 (causing leaf spot on *L. esculentum* and *R. sativus*), a 94% similarity in *gyrB* gene sequence was found relatively to the type strain of *X. vasicola* (the leaf streak pathogen that

affects sorghum). Nevertheless, the phylogenetic distances depicted in Figure 11 are too large to enable an accurate identification of this isolate.

1.3.3 Assessment of partial virulence profiles

According to the presence or absence of the virulence genes *adkXcc*, *dsbB*, *hpaR, hrpF, katG*, *mip-like*, *ppsA* and *zur* a partial virulence profile (PVP) was also established. Overall, five PVP's were determined, as described in Table 7. PVP A, determined by the presence of all eight genes, was recorded for 30 isolates. The remaining 4 PVP's were allocated to one isolate each. PVP B, characterized by the absence of the amplification product of *hpaR* gene, was recorded for CPBF143 isolate. PVP C, recorded for CPBF394, was characterized by the absence of amplification product of *hrpF* gene. Isolate CPBF47 established PVP D, determined by the absence of amplification for *hrpF*, *katG* and *hpaR*. PVP E, established by the type strain CFBP5241^{T} was characterized by the absence of amplification product for *ppsA* gene.

Table 7 – Partial Virulence Profiles (PVP), established for the Portuguese *Xanthomonas campestris* isolates tested in this study, according to the presence/absence of PCR amplification of eight selected virulence genes.

PVP	*adkXcc*	*dsbB*	*hpaR*	*hrpF*	*katG*	*mip-like*	*ppsA*	*zur*
A	+	+	+	+	+	+	+	+
B	+	+	-	+	+	+	+	+
C	+	+	+	-	+	+	+	+
D	+	+	-	-	-	+	+	+
E	+	+	+	+	+	+	-	+

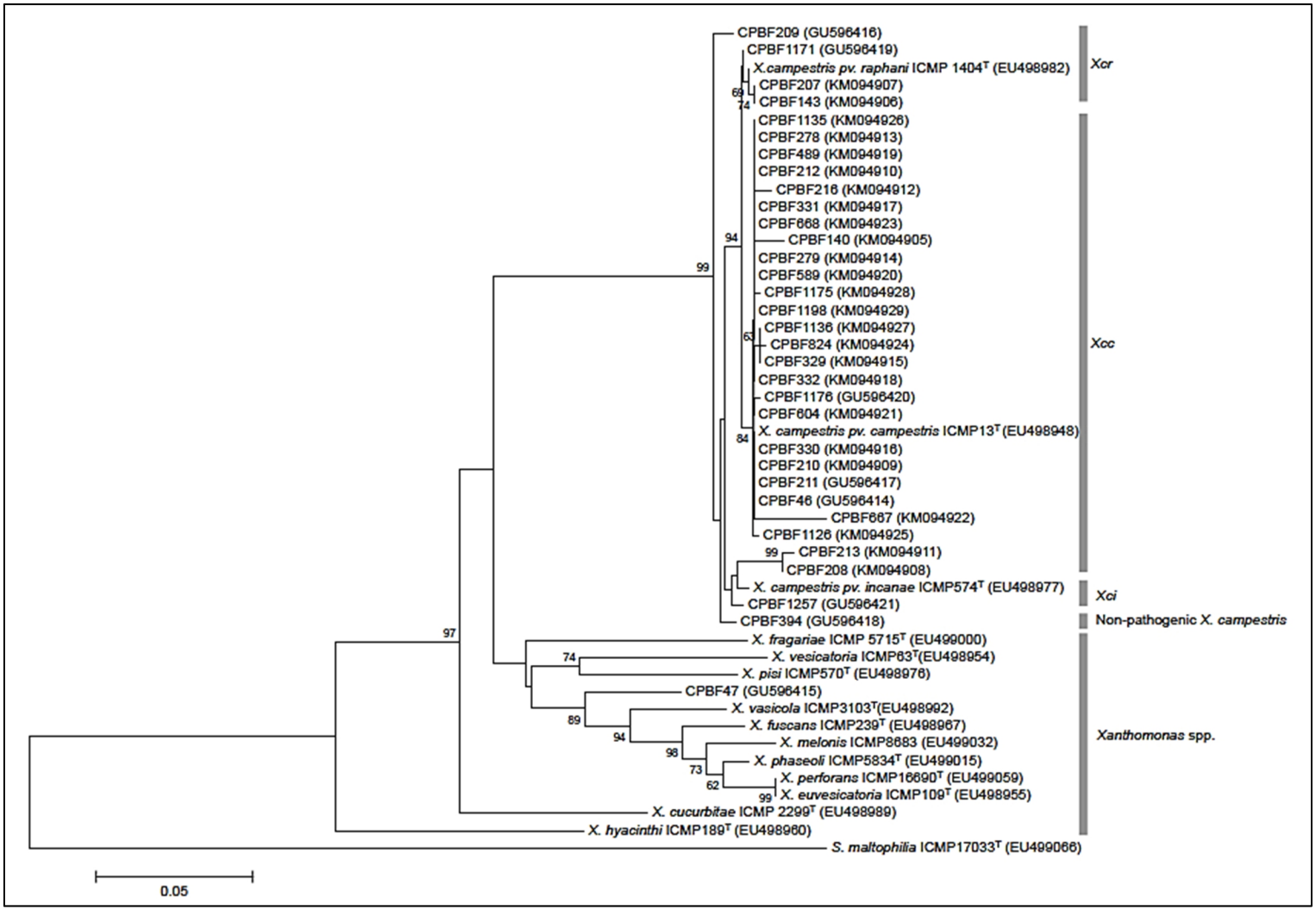


Figure 11 - Neighbor-joining tree based on *gyrB* gene sequences, depicting the phylogenetic relationships among Portuguese *Xanthomonas campestris* isolates and strains of *Xanthomonas* spp. The *gyrB* gene sequence of *Stenotrophomonas maltophilia* ICMP 17033^{T} was used as an outgroup. Percentage bootstrap values >60% from 1000 replications are indicated. Bar: 0.05 substitutions per nucleotide position. Assession numbers are listed in brackets. ^{T}: type strain. *Xcc*: *X. campestris* pv. *campestris*; *Xcr*: *X. campestris* pv. *raphani*; *Xci*: *X. campestris* pv. *incanae*.

1.4 Discussion

Phenotypic characterization and pathogenicity tests, genomic fingerprinting and detection of virulence genes, as well as phylogenetic analysis, showed relevant levels of diversity in the set of isolates used in this study.

Pathovar determination was achieved through the inoculation of *B. oleracea* var. *sabauda* cv. Wirosa, *B. oleracea* var. *costata* cv. 'Beira' and *B. oleracea* var. *acephala* cv. 'Bonanza'. The first cultivar is frequently used to confirm the identity of suspected *Xcc* isolates, as it is considered susceptible to all *Xcc* races. The two latter cultivars are frequently used in Portugal and were included in this study for a complementary comparative characterization of the pathogenicity of the isolates collected in the country. Among the 33 Portuguese isolates studied, 26 isolates causing black rot symptoms on at least one brassicaceous host were identified as *Xcc*. Isolates causing leaf spots on these hosts were identified as *Xcr*. In addition to pathovars *campestris* and *raphani,* the presence of one isolate (CPBF1257) collected from *Erysimum* sp., that showed blight symptoms on *Erysimum* spp. indicates the presence of the third *X. campestris* pathovar – *X. campestris* pv. *incanae*.

Since the establishment of a race structure among *Xcc* in 1992, several changes have been made. The latest revision updated the structure from six to nine races (Fargier & Manceau, 2007), since 10 of the 34 isolates studied by these authors had pathogenicity profiles different with those that had been described earlier (Vicente *et al.*, 2001). Although the relative prevalence of each race in a given location may depend on the host, races 1 and 4 are considered predominant worldwide (Vicente & Holub, 2013). In the present study, *Xcc* race-typing assays performed using the differential host series (Vicente *et al.*, 2001), showed the presence races 4, 6 and 7, encompassing 21 out of 33 isolates. A previous study conducted on a set of 51 isolates collected in Portugal between 1994 and 1997 had already documented the presence of races 1, 4 and 6 (Vicente, 2004). Among those isolates, race 4 was the most prevalent, followed by race 6 and a correlation between races and host of origin was suggested. Race 4 isolates were common in *B. oleracea* varieties but absent in *B. rapa*. According to that study, the majority of race 6 isolates were collected from *B. rapa* and *B. oleracea* vars. *acephala*, *costata* and *capitata* and no isolates were retrieved from *B.oleracea* vars. *italica*, *botrytis* or *sabauda*. On the two latter varieties, race 4 seemed to be prevalent.

Interestingly, in our work, only one isolate, collected from *B. oleracea* var. *acephala*, belonged to race 4. In addition, all race 6 isolates were collected from several *B. oleracea* varieties including var. *botrytis* and *sabauda*. Although race 7 is also represented in Portugal, it is important to notice that this race was only proposed in 2007, by Fargier and Manceau. Overall, our results differ from the ones of Vicente (2004), concerning both the relative prevalence of each race as well as their 'preferred' host

of origin. The collection of race 6 isolates from *B. oleracea* var. *botrytis* and var. *sabauda* could easily be explained by a reintroduction of the pathogen through infected seeds. In the same manner, race 7 may have entered the country after 2003. Two new races, race 10 and race 11, are also reported in this work. Race 10 encompassed three isolates collected from *B. oleracea* (CPBF330, CPBF331 and CPBF604), unable to cause symptoms in *B. rapa* cv. 'Seven Top Turnip' and in *B. juncea* cv. 'Florida Broad Leaf Mustard'. Race 11, represented by isolates CPBF489 and CPBF1198 collected from *B. oleracea* and *B. kaber*, is unable to infect *B. rapa* cv. 'Seven Top Turnip'. These findings contribute to the growing idea that race structure within *Xcc* is dynamic, evolving as new isolates from different geographical origins are studied.

Despite being routinely used, there are still some issues concerning the suitability of the differential host series used for *Xcc* race determination. The current set includes cultivars with variable reactions, such as Seven Top Turnip (Kamoun *et al.*, 1992; Vicente *et al.*, 2001), as well as F1 hybrids, which are impossible for researchers to regenerate. Therefore, the future availability of this host set currently relies on the commercial success of each cultivar, hampering the continuity of the race structure in its current form.

Additionally, variable host responses have also been recorded when different inoculation methods are used (Vicente *et al.*, 2006; Fargier & Manceau, 2007). After spraying inoculation, it is common for the leaves to be covered in small dark spots that eventually evolve to the typical black rot symptoms. On the other hand, the symptoms appear more rapidly and are more obvious after clipping inoculation of the leaves. These differences in symptom development according to the inoculation method may not be important for highly virulent isolates, since after a few days the dark spots evolve to typical V-shaped lesions. However, for low virulence isolates, the two weeks during which plants are checked for symptoms may not be sufficient to allow symptom development and the small dark spots may be confused for other closely related pathogen such as *X. campestris* pv. *raphani*.

The most common methods, spraying and clipping, were also used in this study for pathovar and race identification respectively. Similarly to what was already described, in this study three isolates were only able to cause symptoms in *B. oleracea* cv. 'Wirosa' after clipping inoculation. It should be noted that clipping the leaves with a loaded bacterial suspension is a forceful method of inoculation and may not reflect the natural infection process by which the pathogen enters the host via small wounds or natural openings and is challenged to overcome plant defenses in order to cause disease.

Thus, it seems crucial that the differential host series should be revised in order to include exclusively hosts that show a reliable response to inoculation. Furthermore, we believe that the inoculation method should be fitted in order to mirror the natural infection process. Although time

consuming, the method of race determination currently used would greatly benefit from this adjustment, providing a more realistic assessment of what may be expected under field conditions.

The eight genes screened in this study were shown, through functional characterization, as essential to virulence in *Xcc* isolates. Among the 33 isolates, five different partial virulence profiles (PVP's) were established. For 30 isolates, all genes were present, configuring PVP A, regardless of its pathovar and/or race. This result is in accordance to the fact that all of these isolates were able to cause symptoms in at least one *B. oleracea* variety.

The absence of amplification for *hpaR* target in *Xcr* isolate CPBF143 established PVP B. *hpaR* (hrp [hypersensitive response and pathogenicity]-associated regulator) is a regulatory gene, encoding for a transcriptional regulator of the MarR (multiple antibiotic resistance regulator) family. The MarR members regulate several biological functions of both bacteria and archaea, including resistance to multiple antibiotics, toxic chemicals and oxidative stress agents (Alekshun & Levy, 1999). They are also involved in the regulation of adaptation to different environments and the expression of virulence genes of both plant and animal pathogens (Ellison & Miller, 2006). Mutation of *hpaR* in *Xcc* resulted in complete loss of virulence in *B. oleracea* cv. Jingfeng no.1, delayed and weakened hypersensitive response in the non-host plant *Capsicum anuum* cv. ECW-10R, as well as an increase in extracellular protease production (Wei *et al.*, 2007). Although no previous studies on the role of this gene in *Xcr* have been conducted, its absence may likely result in the inability of this isolate to cause symptoms in any of the *B. oleracea* varieties tested, while other *Xcr* isolates were able to establish a compatible interaction with hosts of this species.

PVP C, characterized by the absence of amplification for *hrpF* gene, was recorded for CPBF394, an isolate obtained from *Ranunculus* sp. that proved to be non-pathogenic for any of the hosts tested. The *hrp* cluster, highly conserved among plant pathogenic bacteria, is essential for host-pathogen interaction, encoding type III secretion systems that deliver virulence factors, elicitors and avirulence proteins to the host cell. The 3' end of *hrpF* was reported to stand on its own as a good target for identifying *X. campestris* pathovars (Berg *et al.*, 2005), with no amplification products from other xanthomonads. Additionally, the absence of this fragment was correlated with the lack of virulence *in planta*. Our results are in accordance to what has been verified by these authors, that also reported two *X. campestris* isolates obtained from *Ranunculus* sp. also lacking the 619 bp fragment of the *hrpF* gene.

The absence of amplification products for *hrpF*, *hpaR* and *katG* genes established PVP D for isolate CPBF47. This isolate, obtained from the brassicaceous host *Rorippa nasturtium-aquaticum,* was not pathogenic in any of the *B. oleracea* tested and therefore should not be included in *X. campestris*. This is in accordance with the absence of amplification for *hrpF*, used for the identification of

X. campestris pathovars but not for other xanthomonads. Additionally, the absence of *hpaR* gene could explain the lack of virulence towards *B. oleracea* hosts, as for the *Xcr* isolate CPBF143 described above. Similarly, the absence of *katG* has also been related to a virulence deficiency of *Xcc* against *R. sativus* (Jittawuttipoka *et al.*, 2009). This gene encodes a catalase-peroxidase that acts as a H_2O_2 scavenger protecting the pathogen's cells against the oxidative burst, which consists of the rapid generation of ROS, primarily H_2O_2, at the site of attempted invasion (Baker & Orlandi, 1995). Although isolate CPBF47 was pathogenic against *R. sativus*, it is important to notice that it was not included in *X. campestris*. Other catalase-peroxidases were found essential for full virulence in other xanthomonads, such as *X. axonopodis* pv. *citri* where the mutation of *katE* led to an attenuate virulence against citrus leaves (Tondo *et al.*, 2010).

Unexpectedly, no successful amplification of *ppsA* gene was obtained for type strain CFBP5241^T, resulting in the establishment of PVP E. This gene encodes for a phosphoenolpyruvate synthase, a key enzyme in gluconeogenesis, which is required for virulence (Tang *et al.*, 2005a). *ppsA* in *Xcc* was described for strain 8004, a spontaneous rifampicin-resistant strain derived from *Xcc* NCPPB1145 (Tang *et al.*, 2005a; Qian *et al.*, 2005). However, the genome of this strain is highly conserved when compared to the type strain and a homolog for *ppsA* in strain CFBP5241^T has been described.

Overall, partial virulence profiles did not correlate with virulence, pathovar, race or origin of the tested isolates. However, for most identified virulence genes the determination of their functional role in virulence was based upon inoculation on a limited host set. Therefore, it would be of great interest if future studies on the functional role of virulence genes will consider pathogenicity assays on a wider host range.

Phylogenetic analysis based on DNA sequencing of constitutive genes has been shown to attain higher resolution power than those based on 16S or 16S-23S intergenic space and, concerning xanthomonads, *gyrB* gene has been used as a preferential target for both identification at species level and phylogenetic studies (Parkinson *et al.*, 2007, 2009). *gyrB*-based phylogenies have clarified some aspects of xanthomonad classification, such as the resolution of *X. citri* and *X. fuscans* which were previously identified as pathovars of *X. axonopodis* by means of DNA-DNA hybridization (Vauterin *et al.*, 1995). This locus was also found helpful in the identification of infra-specific diversity among strains of *X. translucens*, *X. arboricola*, *X. hortorum* and *X. campestris,* since most pathovars of these species were distinguished (Parkinson *et al.*, 2009).

In this study, phylogenetic analysis using *gyrB* gene allowed identifying a high level of diversity among this set of isolates. The topology of the phylogenetic tree constructed on the basis of *gyrB* gene sequencing confirmed the phenotypic identification at species level achieved for the isolates.

Additionally, an overall discrimination of the three described pathovars of *X. campestris* was revealed, as previously indicated by Parkinson *et al.* (2009). However, two isolates described as *Xcc* (CPBF 208 and CPBF 213) grouped together in a separate cluster, along with *Xci* isolate CPBF1257 and the type strain for *Xci*. Despite presenting a higher level of similarity to *Xci* rather than *Xcc*, the identification of isolates CPBF208 and CPBF213 was maintained as *Xcc*, due to the fact that they were originally retrieved from *B. oleracea* plants and caused typical black rot symptoms on at least two of the hosts used for pathovar determination as well as on all the hosts used for *Xcc* race-typing. Analogously, three of the four *Xcr* isolates grouped together with *Xcr* type strain in a cluster separated from *Xcc*. Isolate CPBF209, causing leaf spots on several brassica hosts, appears well separated from all other isolates, thus supporting the previous description of genomic diversity within *Xcr* (Vicente *et al.*, 2006). The existence of variability within *Xcc* was also revealed using this approach, although phylogenetic allocation did not correlate to the race structure defined for this pathovar, nor to the host or geographical origin of these isolates.

In addition to the diversity of isolates described above, two atypical isolates were also characterized. Isolate CPBF 47, originally obtained from the cruciferous plant *Rorippa nasturtium-aquaticum*, caused leaf spots on *S. lycopersicum* and *R. sativus*. It was also able to cause leaf spots on *C. anuum*, *B. carinata* and *B. rapa* cv. 'Just right turnip' (data not shown). This host range determination suggested it to be *Xcr*. However, it was not positive for *hrpF*-based PCR and proved to be quite different from other strains of *X. campestris* by genomic fingerprinting profile. Nucleotide BLAST against available *gyrB* sequences revealed a low similarity level (94%) with isolate ICMP3103^{T}, the pathotype strain for *X. vasicola* pv. *holcicola*, the pathogen responsible for leaf streak of sorghum, a monocotyledonous crop. Further characterization of this isolate is necessary, in order to accurately achieve its identification.

Isolate CPBF 394 also showed distinctive characteristics, albeit being identified as *X. campestris* by cultural and biochemical characterization as well as by phylogenetic analysis. It was negative in PCR with *hrpF* specific primers, not pathogenic on any of the hosts tested, except for *Ranunculus* sp. from which it was originally obtained, and displayed an atypical genomic fingerprinting profile. Further approaches should be undertaken, such as pathogenicity tests on a wider host range and a more exhaustive molecular characterization, through sequencing of additional genes.

The combination of phenotypic, genomic and phylogenetic data throughout this work allowed the characterization of a collection of *X. campestris* isolates collected in Portugal displaying several interesting features. The presence of all *X. campestris* pathovars in Portugal was confirmed, corroborating the recent restriction of *X. campestris* to only three pathovars, causing three distinct diseases. To our knowledge this is the first report of the presence of *X. campestris* pv. *incanae* in

Portugal. The presence of two new *Xcc* race-typing patterns race suggests that the *Xcc* race structure should be revised, in order to better fit the reality of *Xcc* isolates occurring worldwide.

Additionally, the gene-for-gene model based on the interaction of at least five matching gene pairs described in the literature should, ultimately, assist in the assessment of resistant plant material to be used in breeding programs (Vicente & Holub, 2013). However, the lack of genetic and molecular data from both pathogen and host still hinders the full validation of this model, which represents a clear setback on black rot control. The absence of known virulence genes, thought to be crucial for full virulence, in pathogenic isolates is also indicative of the complexity of the pathogenesis process.

This work pointed to several key factors that hamper the establishment of an evident correlation between genomic, phylogenetic and phenotypic features, such as the inoculation method and differential host series used to determine pathovars and races. Furthermore, the use of a polyphasic approach allowed determining the presence of all *X. campestris* pathovars in Portugal and describing two new races among *Xcc*. Despite of the current development of this research field, we can conclude that a detailed study of the interaction established in the pathosystem *X. campestris* - *B. oleracea* is of utmost importance.

References

Alekshun MN, Levy SB, 1999. The *mar* regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends in microbiology* **7**, 410–413.

Baker CJ, Orlandi EW, 1995. Active oxygen in plant pathogenesis. *Annual review of phytopathology* **33**, 299–321.

Berg T, Tesoriero L, Hailstones DL, 2005. PCR-based detection of *Xanthomonas campestris* pathovars in Brassica seed. *Plant Pathology* **54**, 416–427.

Bogdanove AJ, Koebnik R, Lu H *et al.*, 2011. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *Journal of bacteriology* **193**, 5450–5464.

Bolot S, Guy E, Carrere S, Barbe V, Arlat M, Noël LD, 2013a. Genome Sequence of *Xanthomonas campestris* pv. *campestris* Strain Xca5. *Genome announcements* **1**, e00032-12.

Bolot S, Roux B, Carrere S *et al.*, 2013b. Genome sequences of three atypical *Xanthomonas campestris* pv. *campestris* strains, CN14, CN15, and CN16. *Genome announcements* **1**, e00465-13.

Cruz J, 2009. Diversidade genómica, patogenicidade e sensibilidade a compostos bioactivos em *Xanthomonas campestris* . Faculdade de Ciências da Universidade de Lisboa.

Ellison DW, Miller VL, 2006. Regulation of virulence by members of the MarR/SlyA family. *Current opinion in microbiology* **9**, 153–159.

Fargier E, Manceau C, 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**, 805–818.

He Y-Q, Zhang L, Jiang B-L *et al.*, 2007. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome biology* **8**, R218.

Jiang B-L, Liu J, Chen L-F *et al.*, 2008. DsbB is required for the pathogenesis process of *Xanthomonas campestris* pv. *campestris*. *Molecular plant-microbe interactions : MPMI* **21**, 1036–1045.

Jittawuttipoka T, Buranajitpakorn S, Vattanaviboon P, Mongkolsuk S, 2009. The catalase-peroxidase KatG is required for virulence of *Xanthomonas campestris* pv. *campestris* in a host plant by providing protection against low levels of H2O2. *Journal of bacteriology* **191**, 7372–7377.

Kamoun S, Kamdar H V, Tola E, Kado CI, 1992. Incompatible Interactions Between Crucifers and

Xanthomonas campestris Involve a Vascular Hypersensitive Response : Role of the *hrpX* Locus. *Molecular Plant-Microbe Interactions* **5**, 22–33.

Lu GT, Tang YQ, Li CY *et al.*, 2009. An adenosine kinase exists in *Xanthomonas campestris* pathovar *campestris* and is involved in extracellular polysaccharide production, cell motility, and virulence. *Journal of Bacteriology* **191**, 3639–3648.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D, 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **57**, 2881–2887.

Parkinson N, Cowie C, Heeney J, Stead D, 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *International Journal of Systematic and Evolutionary Microbiology* **59**, 264–274.

Pinto-Ganhão J, 1962. *Xanthomonas campestris* (Pammel) Dowson - Uma nova bacteriose para Portugal. *Agros* **45**, 167–168.

Qian W, Jia Y, Ren S-X *et al.*, 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome research* **15**, 757–767.

da Silva ACR, Ferro JA, Reinach FC *et al.*, 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA 5: Molecular Evolutionary Genetics Analysis using Maximum Likelyhood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731–2739.

Tang D, He Y, Feng J *et al.*, 2005a. *Xanthomonas campestris* pv. *campestris* possesses a single gluconeogenic pathway that is required for virulence. *Journal of bacteriology* **187**, 6231–6237.

Tang D-J, Li X-J, He Y-Q, Feng J-X, Chen B, Tang J-L, 2005b. The zinc uptake regulator Zur is essential for the full virulence of *Xanthomonas campestris* pv. *campestris*. *Molecular plant-microbe interactions : MPMI* **18**, 652–658.

Tondo ML, Petrocelli S, Ottado J, Orellano EG, 2010. The monofunctional catalase KatE of *Xanthomonas axonopodis* pv. *citri* is required for full virulence in citrus plants. *PloS one* **5**, e10803.

Vauterin L, Hoste B, Kersters K, Swings J, 1995. Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology* **45**, 472–489.

Vicente JG, 2004. *A Podridão Negra das Crucíferas* (G Lopes, Ed,). Alcobaça: COTHN Centro Operativo Tecnológico Hortofruticola.

Vicente JG, Conway J, Roberts SJ, Taylor JD, 2001. Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathovars. *Phytopathology* **91**, 492–499.

Vicente JG, Everett B, Roberts SJ, 2006. Identification of isolates that cause a leaf spot disease of Brassicas as *Xanthomonas campestris* pv. *raphani* and pathogenic and genetic comparison with related pathovars. *Phytopathology* **96**, 735–745.

Vicente JG, Holub EB, 2013. *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular plant pathology* **14**, 2–18.

Vicente JG, Taylor JD, Sharpe a G, Parkin I a P, Lydiate DJ, King GJ, 2002. Inheritance of Race-Specific Resistance to *Xanthomonas campestris* pv. *campestris* in Brassica Genomes. *Phytopathology* **92**, 1134–1341.

Vorhölter F-J, Schneiker S, Goesmann A *et al.*, 2008. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *Journal of biotechnology* **134**, 33–45.

Wei K, Tang D-J, He Y-Q *et al.*, 2007. hpaR, a putative marR family transcriptional regulator, is positively controlled by HrpG and HrpX and involved in the pathogenesis, hypersensitive response, and extracellular protease production of Xanthomonas campestris pathovar campestris. *Journal of bacteriology* **189**, 2055–2062.

Zang N, Tang D-J, Wei M-L *et al.*, 2007. Requirement of a mip-like gene for virulence in the phytopathogenic bacterium *Xanthomonas campestris* pv. *campestris*. *Molecular plant-microbe interactions : MPMI* **20**, 21–30.

2. Inference of phylogenetic diversity and population structure of *Xanthomonas campestris* affecting Brassicaceae using a MLST-based approach

2.1. Introduction

Xanthomonas campestris is currently divided into three pathovars, causing three distinct diseases: i) black rot of cruciferous plants caused by *X. campestris* pv. *campestris* (*Xcc*); ii) leaf spot disease of Brassicaceae and Solanaceae hosts, caused by *X. campestris* pv. *raphani* (*Xcr*); and iii) bacterial blight of ornamental crucifers, caused by *X. campestris* pv. *incanae* (*Xci*) (Fargier & Manceau, 2007). Further differentiation into races has been established for *Xcc* and *Xcr* strains, due to the presence of different symptoms on differential host series. *Xcc* has been divided into nine races (Fargier & Manceau, 2007) and *Xcr* into three races (Vicente *et al.*, 2006). Although no race structure has been adopted for *Xci*, some host specificity has been described (Vicente *et al.*, 2006).

Following this intraspecific differentiation, questions concerning the genetic relationships among and within *X. campestris* pathovars have been brought to light. Traditional phylogenetic analyses targeting 16S rRNA and *gyrB* (DNA gyrase subunit B) genes have failed to properly address this issue. While 16S rRNA gene has shown a discriminatory potential at genus level (Hauben *et al.*, 1997), *gyrB* gene only allows to distinguish strains of *X. campestris* from other species and its use is unfit to clearly differentiate *X. campestris* pathovars (Parkinson *et al.*, 2009).

Multi Locus Sequence Typing (MLST) is a nucleotide-based approach for the characterization of bacterial pathogens that allows to assess genomic relatedness and population structure as well as to get insights into their evolution, constituting a valuable tool for global epidemiology (Urwin & Maiden, 2003). The analysis of the data generated by MLST has been successfully used to characterize evolution and population genetics of several plant-pathogenic bacteria, such as *Pseudomonas syringae* (Sarkar & Guttman, 2004), *Ralstonia solanacearum* (Castillo & Greenberg, 2007) and *Xylella fastidiosa* (Yuan *et al.*, 2010). MLST has also been applied to *X. campestris* in an attempt to determine the relationships between phylogeny and phenotypic characteristics of a collection of isolates (Fargier *et al.*, 2011). In that study, the phylogenetic and split decomposition analysis of concatenated sequences for housekeeping genes revealed an association of genetic lineages to specific *X. campestris* pathovars, suggesting that the genetic structure of the species has been shaped by the differential adaptation to host plants. The application of eBURST algorithm to MLST data also allowed the identification of the putative ancestral genotype of *X. campestris* that surprisingly was not included in any of the established clonal complexes.

More recently, new approaches have been developed for the analysis of MLST data, replacing nucleotide-based for allelic-based analyses. Through the use of goeBURST, a globally optimized implementation of the eBURST algorithm (Feil *et al.*, 2004), a Minimum Spanning (MS) tree revealing the possible evolutionary relationships can be constructed (Francisco *et al.*, 2009). When associated with a user-friendly software, this new approach allows the overlapping of MLST data with other data sources in a combined analysis for epidemiological and population studies (Francisco *et al.*, 2012). Using this method, the founder genotype, defined by the MLST data, appears in a central position in the population, from which the remaining genotypes derive. This graphic matroid approach was recently applied to the study of *Klebsiella pneumoniae* and other relevant human bacterial pathogens (Guo *et al.*, 2015; Abdulgader *et al.*, 2015).

In the previous chapter, for a set of Portuguese isolates, the presence of all three pathovars of *X. campestris* was recorded and the race structure among *Xcc* was also corroborated with the description of two new races. Additionally, the high level of phenotypic, genomic and phylogenetic diversity among Portuguese *X. campestris* isolates was also revealed. In this chapter, an integrative analysis of MLST and phenotypic data, using Neighbor-Joining phylogenetic analysis and goeBURST algorithm was implemented to characterize 75 *X. campestris* isolates from distinct geographical origins, pathovars and race profiles, to disclose their populational structure and genetic diversity as well as the underlying phylogenetic and epidemiological patterns.

2.2. Materials and Methods

2.2.1. Bacterial isolates

A collection of 75 isolates was used for MLST analysis, as described in Table 8. The subset of 33 isolates from the Portuguese Collection of Phytopathogenic Bacteria (CPBF) includes 26 *Xcc* isolates, four *Xcr* isolates, one *Xci* isolate and two atypical isolates previously identified as *X. campestris*. Before use, all isolates were grown for 48h at 27°C in Yeast Dextrose Chalk Agar media (Lelliot & Stead, 1987).

Table 8 - *Xanthomonas campestris* isolates used in this study, their pathovar and race classification, sequence type and allelic profiles considering the four loci *dnaK*, *fyuA*, *gyrB* and *rpoD*.

Isolate number	Race[a]	Isolation			Sequence Type	Allelic Profiles			
		Host	Year	Country		*dnaK*	*fyuA*	*gyrB*	*rpoD*
Xanthomonas campestris pv. *campestris*									
Xcc 147	4	*B. oleracea*			ST1	1	1	1	1
CPBF 213	6	*B. oleracea* var. *costata*	2004	Portugal	ST2	1	3	1	1
CPBF 1126	6	*B. oleracea*	2002	Portugal	ST2	1	3	1	1
CPBF 331	10	*B. oleracea* var. *costata*	2004	Portugal	ST2	1	3	1	1
CFBP 4956	4	*B. oleracea* var. *botrytis*	1999	Belgium	ST3	1	3	1	4
CFBP 6650	9	*B. oleracea*	1958	UK	ST3	1	3	1	4
CFBP 5820	8	*B. oleracea*			ST3	1	3	1	4
CFBP 1124	8	*B. oleracea* var. *botrytis*	1967	France	ST4	1	3	1	9
CFBP 1710	8	*B. oleracea* var. *botrytis*	1975	France	ST4	1	3	1	9
CFBP 1711	8	*B. oleracea* var. *botrytis*	1975	France	ST4	1	3	1	9
CFBP 6863	8	*B. oleracea* var. *botrytis*	1958	Germany	ST4	1	3	1	9

Isolate number	Race[a]	Isolation			Sequence Type	Allelic Profiles			
		Host	Year	Country		*dnaK*	*fyuA*	*gyrB*	*rpoD*
CPBF 279	6	*B. oleracea*		Portugal	ST5	1	3	1	11
CPBF 1135	6	*B. oleracea* convar. *capitata* var. *alba*	2002	Portugal	ST5	1	3	1	11
CFBP 4952	1	*B. oleracea* var. *botrytis*	1999	Belgium	ST5	1	3	1	11
CFBP 4953	7	*B. oleracea* var. *botrytis*	1999	Belgium	ST5	1	3	1	11
CFBP 1119	1	*B. oleracea* var. *botrytis*	1967	France	ST5	1	3	1	11
CFBP 1869	1	*B. oleracea*	1976	Ivory Coast	ST5	1	3	1	11
CFBP 5815	1	*B. oleracea* var. *botrytis*	2000	Spain	ST5	1	3	1	11
CFBP 5127	1	*B. oleracea*	2000		ST5	1	3	1	11
CFBP 5128	1	*B. oleracea*	2000		ST5	1	3	1	11
CFBP 5129	1	*B. oleracea*	2000		ST5	1	3	1	11
CFBP 5130	7	*B. oleracea*	2000		ST5	1	3	1	11
CPBF 46	4	*B. oleracea* convar. *acephala* var. *sabellica*	2003	Portugal	ST6	1	3	1	14
CPBF 140	7	*B. oleracea* var. *costata*	2003	Portugal	ST7	1	3	1	16
CPBF 278	6	*B. oleracea*	1992	Portugal	ST8	1	3	1	21
CFBP 6865	5	*B. oleracea* var. *capitata*	1975	Australia	ST9	1	3	3	11
CPBF 667	6	*B. oleracea*		Portugal	ST10	1	4	1	4
CPBF 1136	6	*B. oleracea* convar. *capitata* var. *alba*	2002	Portugal	ST10	1	4	1	4
CPBF 210	7	*B. oleracea* convar. *botrytis* var. *botrytis*	2004	Portugal	ST11	1	4	1	11
CPBF 1175	6	*B. oleracea* convar. *botrytis* var. *botrytis*	2006	Portugal	ST12	1	4	1	12
CFBP 5817	4	*B. oleracea* var. *botrytis*	2001	Chile	ST12	1	4	1	12
CFBP 5816	1	*B. oleracea* var. *botrytis*	2000	Spain	ST12	1	4	1	12
CFBP 5818	1	*B. oleracea* var. *botrytis*	2001	Spain	ST12	1	4	1	12
CPBF 329	7	*B. oleracea* var. *costata*	2004	Portugal	ST13	1	4	1	22
CPBF 1176	6	*B. oleracea* convar. *botrytis* var. *botrytis*	2006	Portugal	ST14	1	4	1	31
CFBP 3838	NP	*Iberis* sp.	1954	Tanzania	ST15	1	10	1	12
CPBF 824	7	*B. oleracea*		Portugal	ST16	1	23	1	29
CFBP 5814	9	*B. oleracea* var. *botrytis*	2000	Spain	ST18	3	2	6	7
CFBP 5241^T	3	*B. oleracea* var. *gemmifera*	1957	UK	ST25	9	3	1	4
CFBP 4954	6	*B. oleracea* var. *botrytis*	1999	Belgium	ST26	9	3	1	11
CFBP 4955	9	*B. oleracea* var. *botrytis*	1999	Belgium	ST26	9	3	1	11
CFBP 1121	1	*B. oleracea* cv. *bullata gemmifera*	1967	France	ST26	9	3	1	11
CFBP 1712	5	*B. oleracea* var. *capitata*	1975	France	ST27	9	13	1	11
CFBP 1713	5	*B. oleracea* var. *botrytis*	1975	France	ST27	9	13	1	11
CFBP 5683	3	*B. oleracea*	1979	France	ST27	9	13	1	11
CPBF 1198	11	*B. kaber* var. *pinnatifida*	2007	Portugal	ST30	11	3	1	4
CPBF 208	6	*B. oleracea* var. *costata*	2003	Portugal	ST34	15	18	10	19
CPBF 211	7	*B. oleracea* convar. *botrytis* var. *botrytis*	2004	Portugal	ST36	17	4	1	6
CPBF 212	7	*B. oleracea var. costata*	2004	Portugal	ST37	18	3	12	12
CPBF 216	7	*B. oleracea* convar. *botrytis* var. *botrytis*	2005	Portugal	ST38	19	20	1	20
CPBF 330	10	*B. oleracea* var. *costata*	2004	Portugal	ST39	20	3	1	23
CPBF 332	7	*B. oleracea* var. *costata*	2004	Portugal	ST40	21	3	1	24
CPBF 489	11	*B. oleracea* var. *costata*	1997	Portugal	ST43	24	4	14	26
CPBF 589	6	*B. oleracea* convar. *botrytis* var. *botrytis*	2005	Portugal	ST44	25	3	1	27
CPBF 604	10	*B. oleracea* convar. *capitata* var. *sabauda*	2005	Portugal	ST45	26	3	1	1
CPBF 668	7	*B. oleracea*	1998	Portugal	ST46	27	22	1	28
Xanthomonas campestris pv. *raphani*									
CFBP 5827	NA	*Raphanus sativus*	1940	USA	ST17	2	12	1	5
CFBP 5829	NA	*Raphanus sativus*		USA	ST19	4	7	1	10
CFBP 5828	NA	*Raphanus sativus*		USA	ST20	5	11	1	6
F756C	NA	*B. oleracea* var. *capitata*			ST28	10	15	1	13
CPBF 1171	NA	*B. oleracea* convar. *capitata* var. *sabauda*	2006	Portugal	ST29	10	24	1	30
CPBF 143	NA	*B. oleracea* convar. *capitata* var. *alba*	2003	Portugal	ST32	13	16	8	17
CPBF 207	NA	*B. oleracea* convar. *botrytis* var. *botrytis*	2003	Portugal	ST33	14	17	9	18
CPBF 209	NA	*B. oleracea* var. *costata*	2004	Portugal	ST35	16	19	11	9
Xanthomonas campestris pv. *incanae*									

Isolate number	Race[a]	Isolation			Sequence	Allelic Profiles			
		Host	Year	Country	Type	*dnaK*	*fyuA*	*gyrB*	*rpoD*
CFBP 1371	NA	*Matthiola incana*	1972	France	ST23	8	5	2	3
CFBP 5686	NA	*Matthiola incana*	1981	France	ST23	8	5	2	3
CFBP 1438	NA	*Matthiola incana*	1949	USA	ST23	8	5	2	3
CFBP 2527	NA	*Matthiola incana*	1950	USA	ST23	8	5	2	3
CFBP 1606	NA	*Matthiola incana*	1974	France	ST24	8	8	5	2
CPBF 1257	NA	*Erysimum* sp.	2008	Portugal	ST31	12	8	15	11
Xanthomonas campestris									
CFBP 5825	NA	*Barbarea vulgaris*	1939	USA	ST21	6	6	4	8
CFBP 5826	NA	*Barbarea vulgaris*	1939	USA	ST21	6	6	4	8
CFBP 5824	NA	*Armoracia lapathifolia*	1939	USA	ST22	7	9	2	8
CPBF 394	NP	*Ranunculus* sp.	2005	Portugal	ST41	22	21	13	25
Xanthomonas sp.									
CPBF 47	NP	*Rorippa nasturtium-aquaticum*	2003	Portugal	ST42	23	14	7	15

[a] NP: non-pathogenic; NA: not applicable

2.2.2. MLST scheme

MLST analysis was performed using a modified version of the MLST scheme previously described for *X. campestris* that retained the congruent *dnaK* (heat shock protein 70, molecular chaperone DnaK), *fyuA* (tonB-dependent receptor), *rpoD* (RNA polymerase sigma-70 factor) and *gyrB* loci (Fargier *et al.*, 2011). For the analysis, available sequences were retrieved from GenBank and amplification of missing sequences was carried out as described below.

DNA sequences from each of the four MLST loci that differed from each other by one or more polymorphisms were assigned with different allele numbers. Distinct allelic profiles were assigned distinct sequence types (STs). Two different STs sharing three of the four loci constituted a single-locus variant (SLV). A double-locus variant (DLV) contained two STs differing in two loci. A clonal complex was determined as being composed of at least three STs with only SLVs. The remaining STs, sharing no more than one locus, were termed singletons.

2.2.3. DNA isolation, PCR amplification and Sequencing

For DNA extraction, a water suspension from each bacterial isolate (OD600=0.1) was heated during 7 min at 96°C and then stored at -20°C. *dnaK*, *rpoD* and *fyuA* genes were amplified and sequenced using previously described primers (Fargier *et al.*, 2011). Amplifications were carried out in a final volume of 20 µL containing 1x PCR Buffer (Invitrogen), 200 µM of each dNTP (Invitrogen), 400 nM of each primer, 1.5 mM $MgCl_2$, 1 unit of Taq DNA Polymerase (Invitrogen) and 2 µL of genomic DNA extract. Reactions were performed with initial denaturation at 94°C for 3 min, 35 cycles of 50s at 94°C, 50 s at appropriate annealing temperature (60°C for *gyrB* and 62°C for *dnaK*, *fyuA* and *rpoD*) and 1 min at 72°C, and followed by final extension of 7 min at 72°C. PCR products were then purified using JetQuick PCR Purification Spin Kit (GENOMED, Löhne, Germany). Sequencing reactions were performed on both forward and reverse directions, using an automatic sequencer CEQ 2000-XL

(Beckman Coulter, USA). The assession numbers of the sequences used in this study are described in Table 9. Sequences were initially analysed and corrected using Chromas Lite 2.4.4 software (www.technelysium.com.au) and compared by nucleotide BLAST with those available at GenBank (NCBI). Nucleotide sequences were aligned with MEGA 5.1 software, using default parameters. Both ends of each alignment were trimmed to 527 positions for *dnaK*, 554 positions for *fyuA*, 521 positions for *gyrB* and 665 positions for *rpoD*. Sequences were also concatenated to give a total length of 2267 positions.

Table 9 - Assession numbers and references of sequences of *Xanthomonas campestris* isolates analysed in this study.

Locus	Assession[a]	Reference
dnaK	EF673149 - EF673158	Fargier *et al.* (2011)
	KM272636 - KM272668	This study
fyuA	EF673166 - EF673180	Fargier *et al.* (2011)
	KM272669 - KM272701	This study
gyrB	EF673192 - EF673197	Fargier *et al.* (2011)
	KM094906 - 07; GU596416; GU596419	Cruz *et al.* (2015)
	GU596414-15; GU596417-18; GU596420-21; KM094905; KM094908-29	Cruz *et al.* (2017)
rpoD	EF673198 - EF673210	Fargier *et al.* (2011)
	KM272702 - KM272734	This study

[a] Assession numbers refer to individual isolates, except for those from Fargier *et al.* (2011) that correspond to unique allelic profiles for each locus.

2.2.4. Sequence diversity and recombination analyses

Sequence analysis was performed using several features of DnaSP 4.0 and START2 package. In order to provide information about the degree of selection operating on each gene, the ratio (d_N/d_S) of the non-synonymous (d_N) to the synonymous substitutions (d_S) was investigated, using the Nei and Gojobori method (Nei & Gojobori, 1986). A ratio greater than one denotes positive selection; less than one implies purifying selection; and a ratio of one indicates neutral selection. The Index of Association (I_A) was also determined, to measure the extent of linkage equilibrium within a population by quantifying the amount of recombination among a set of sequences and detecting association between alleles at different loci, since frequent recombination events induce linkage equilibrium (Maynard-Smith *et al.*, 1993).

2.2.5. Population structure and phylogenetic analyses

Phylogenetic analyses were performed using MEGA 5.1 software, both on individual gene sequences and on the concatenated sequences. Neighbor-joining (NJ) trees were generated using the Jukes-Cantor distances method and 1000 bootstrap replicates for all sequences. Due to the known complexity of the phylogenetic relatedness amongst *X. campestris*, a neighbour-net (N-net) was constructed based on the concatenated data for the four loci, using SplitsTree V.4.13.1 (Huson & Bryant, 2006). Global optimal eBURST algorithm implemented using Phyloviz software (Francisco *et*

al., 2012) was used to cluster STs with single-locus variant (SLV) limitation, generating a Minimum Spanning (MS) tree to visualize possible evolutionary relationships between STs.

2.3. Results

Sequence alignment of each of the four loci showed no insertion/deletion events and the concatenated sequence was 2267 nt in length. All loci were polymorphic and the number of variable sites ranged from 54 (10%) to 122 (18%) for *dnaK* and *rpoD*, respectively (Table 10). The nucleotide diversity per site ranged from 0.0076 (*dnaK*) to 0.0362 (*rpoD*). The d_N/d_S ratio for the four loci ranged from 0.0623 (*dnaK*) to 0.1317 (*rpoD*). The value of the I_A for concatenated sequences was 0.9094 (P=0.000). Based on the existence of at least one polymorphism, the number of identified alleles at each locus ranged from 15 (*gyrB*) to 31 (*rpoD*). The number of unique alleles, represented by a single isolate, ranged from12 (*gyrB*) to 23 (*rpoD*). The combination of allele numbers into allelic profiles resulted in the establishment of 46 STs. 20 STs constituting single-locus variants (SLVs) were assigned into one single clonal complex (CC1), while the remaining ones were termed singletons. ST5, comprising 11 isolates, was the most frequent ST and is thus considered the founder ST.

Table 10 - Sequence variation for the four genes analysed by MLST, using sequence data from 75 *Xanthomonas campestris* isolates analysed in this study.

Locus	*dnaK*	*fyuA*	*gyrB*	*rpoD*
Fragment size (bp)	527	554	521	665
No. of alleles	27	24	15	31
No. of unique alelles	22	18	12	23
No. of polymorphic sites (%)	54 (10.2)	90 (16.2)	88 (16.9)	122 (18.3)
Nucleotide diversity per site	0.0076	0.0117	0.0088	0.0362
Average d_N/d_S ratio[a]	0.0623	0.0947	0.0858	0.1317

[a] d_N nonsynonymous substitutions ; d_S synonymous substitutions

The NJ tree constructed with the concatenated sequences from the 46 STs showed that *X. campestris* forms a monophyletic group, distinct from other *Xanthomonas* species and subdivided into five major clusters (Figure 12A). Cluster Xcc1, comprising 12 STs, includes the founder ST5 and ST31, corresponding to the single *Xci* isolate collected in Portugal from *Erysimum* sp. (CPBF1257). While Xcc2 included 9 STs, cluster Xcc3, grouping 6 STs, included ST25, corresponding to isolate CFBP 5241, the type strain for the species. Xci cluster grouped together *Xci* isolates collected from ornamental crucifers, as well as a weakly pathogenic *Xcc* isolate (ST18 - CFBP5814), the *Xcr* pathotype strain (ST17 – CFBP5827) and non-pathogenic isolates (ST21 – CFBP5825, CFBP5826; ST22 – CFBP5824). In addition to 6 *Xcr* STs, *Xcr* cluster included ST4 and ST36, corresponding to *Xcc* isolates collected from cauliflower, as well as a non-pathogenic isolate (ST41 – CPBF394). *Xcc* isolate CPBF208 (ST34), *Xcr* isolate CPBF143 (ST32) and *Xanthomonas* sp. isolate CPBF47 (ST42) were not allocated to any specific cluster. The overall topology of the tree using the concatenated dataset is fully congruent with those obtained from each individual locus (data not shown).

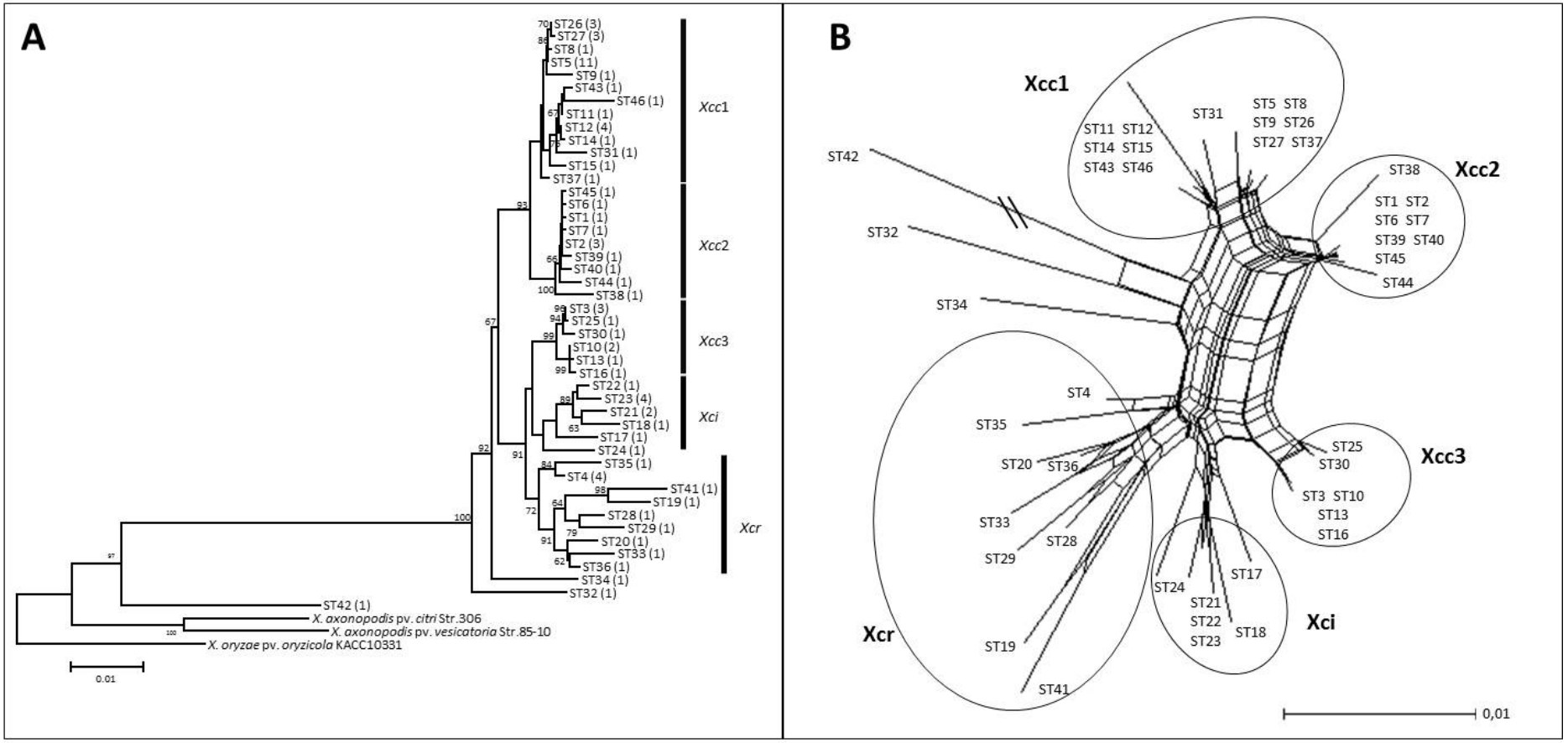


Figure 12 - Diagrams denoting population structure. (a) Neighbour-Joining tree of the 46 sequence types (ST) based on the concatenated sequences of *dnaK*, *fyuA*, *gyrB* and *rpoD*. The number of isolates grouped in each ST is given in brackets. Bootstrap scores greater than 60 are displayed. Major clusters are highlighted. (b) Neighbour-net of the 46 STs. The designation at the branches indicates the ST number. The major clusters observed in the Neighbour-Joining tree are highlighted. For isolate designation and ST correspondence see Table 8. *Xcc*: *Xanthomonas campestris* pv. *campestris*; *Xcr*: *Xanthomonas campestris* pv. *raphani*; *Xci*: *Xanthomonas campestris* pv. *incanae*.

The N-net constructed with the concatenated dataset for the 46 STs revealed an intricate structure, with a complexity of parallelograms, both among and within *X. campestris* pathovars (Figure 12B), suggesting the possibility of recombination events. The clusters defined by the NJ phylogenetic analysis are also depicted in the N-net.

Portuguese isolates were evenly distributed both in the NJ tree and the Neighbour-Net, reflecting the topology of the NJ phylogenetic tree and N-net constructed using the concatenated dataset only for the Portuguese isolates (Appendix).

The goeBURST analysis of the allelic profiles of all 75 isolates generated a minimum spanning (MS) tree (Figure 13), where SLVs and DLVs between STs are displayed. The host group of origin, geographical origin and *Xcc* races of the isolates that constitute each ST are also highlighted.

The MS tree revealed a population structure clearly separating *X. campestris* pathovars, as no ST comprised isolates from different pathovars. The 56 *Xcc* isolates were represented by 30 STs, 20 of which were included into a Clonal Complex (CC1). CC1 is visible as a complex sub-tree of STs deriving from the founder genotype ST5, represented by 11 isolates. *Xcr* comprised 8 STs, all singletons. Among *Xcr* isolates relationships are rare, evidenced by the sole existence of a DLV between ST28 and ST29. Interestingly, ST36 (*Xcc* isolate CPBF211 collected from cauliflower) is simultaneously a DLV of *Xcc* ST11 and *Xcr* ST20 and corresponds to the only visible link between STs of these two pathovars. The 4 *Xci* isolates are represented by 3 apparently unrelated STs.

The correspondence of ST data and the host group of origin is depicted in Figure 13a. For simplification purposes, original host species were separated into major groups. 'Inflorescence kale' includes *B. oleracea* var. *botrytis* and *B. oleracea* var. *italica*; 'kale' group includes *B. oleracea* var. *costata* and *B. oleracea* var. *acephala*; 'cabbage' group includes *B. oleracea* var. *capitata* and *B. oleracea* var. *gemmifera*; '*B. oleracea*' group includes isolates for which no varietal information was available; 'ornamental Brassicaceae' group includes *Mathiola* sp., *Erysimum* sp. and *Iberis* sp.; 'other Brassicaceae crops' group includes *Raphanus sativus* and *Armoracia lapathifolia*; 'Spontaneous Brassicaceae' includes *Barbarea vulgaris* and *Brassica kaber*; isolates from non-Brassicaceae hosts and with unavailable information were grouped in 'other/unknown'. There was no clear correlation between ST attribution and host group of origin, as highlighted by the presence of isolates collected from different hosts in several STs represented by more than one isolate. For instance, the founder ST5 is composed by *Xcc* isolates collected from inflorescence kales, kales and unknown varieties of *B. oleracea*. However, adherence of *Xcc* and *Xcr* STs to Brassicaceae crops and of *Xci* to ornamental Brassicaceae is evident.

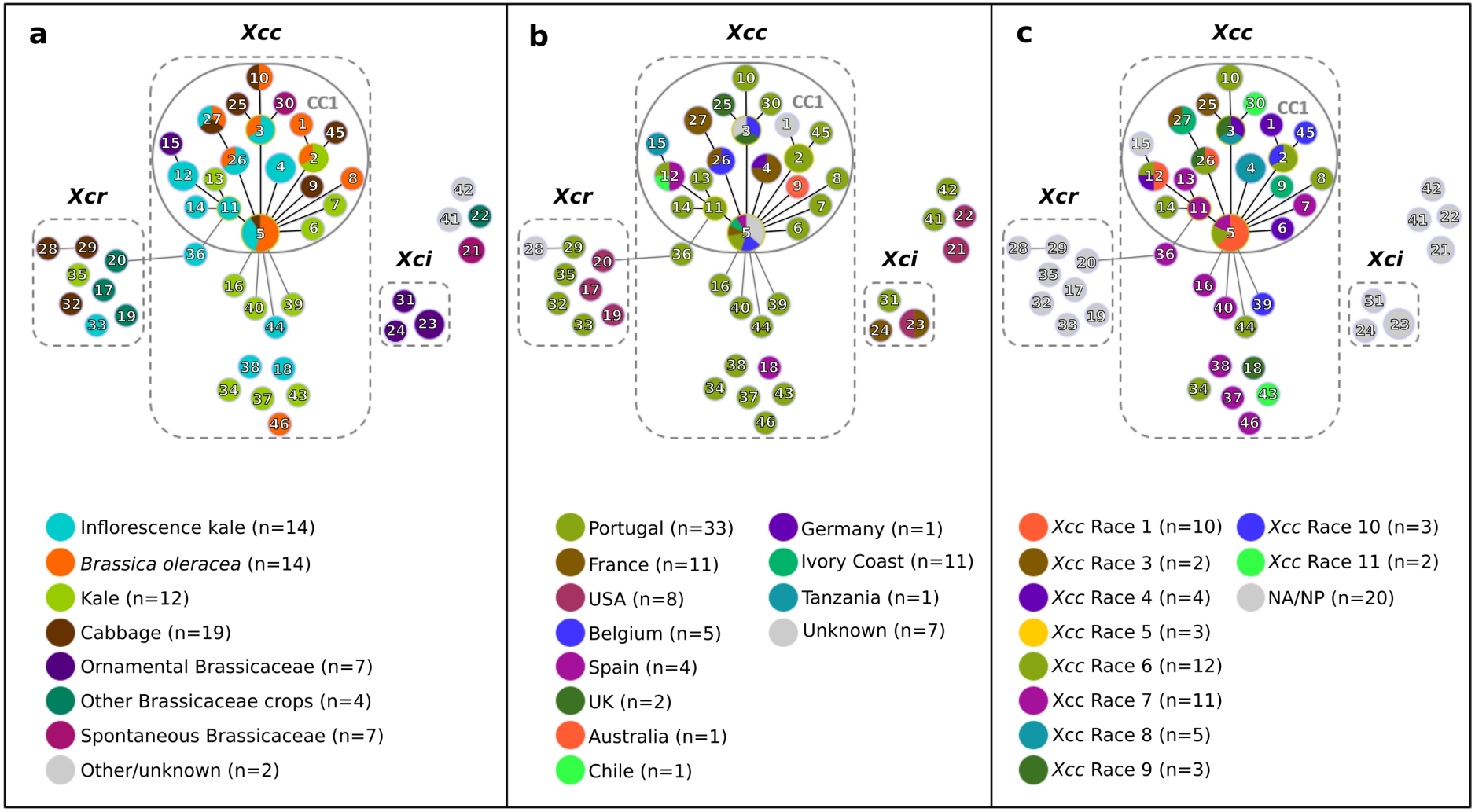


Figure 13 - Minimum Spanning tree of the 46 STs, inferred by Phyloviz based on allelic profiles. Each circle (node) indicates a Sequence Type (ST), and the size of the circle corresponds to the number of isolates included. The edge between Sequence Types represents the relationships of Single Locus Variant (black and solid) and Double Locus Variant (gray and solid). The dashed lines separate three pathovars: *Xcc* (*Xanthomonas campestris* pv. *campestris*), *Xcr* (*Xanthomonas campestris* pv. *raphani*) and *Xci* (*Xanthomonas campestris* pv. *incanae*). The solid gray line grouping *Xcc* STs corresponds to Clonal Complex 1 (CC1). The host group of origin (a), geographical origin (b) and Xcc races (c) of the isolates that constitute each ST are highlighted with different colours. n, number of isolates. NA/NP, not applicable/non-pathogenic.

Similarly, there is no overlapping of ST attribution and geographical origin of the isolates that constitute them (Figure 13b). However, it is important to mention that 26 out of the 33 Portuguese isolates corresponded to unique STs.

The distribution of *Xcc* races among this pathovar also appears to be independent of ST attribution (Figure 13c), except for ST4, which is exclusively represented by 4 isolates belonging to race 8. The founder ST5 includes isolates belonging to races 1, 6 and 7. Races 10 and 11, recorded only in Portugal, also include isolates from different STs.

2.4. Discussion

In this study, a worldwide population of *X. campestris* isolates was studied using MLST, revealing the phylogenetic relationships between isolates and population structure.

The level of polymorphism of the four analysed loci was clearly higher than what was previously described for *X. campestris* (Fargier *et al.*, 2011), suggesting that there was a significant input of genetic diversity associated with the inclusion of the Portuguese isolates in the studied population. In accordance to the MLST requirements and as revealed by d_N/d_S ratios < 1.0, the loci were not under any unusual selective pressure, confirming the absence of diversifying selection in these genes and excluding their direct involvement in host-pathogen interactions (Doyle & Gaut, 2000).

Previous phylogenetic analysis of Portuguese isolates (data not shown) allowed differentiating *Xcc* isolates into three distinct clusters and *Xcr* isolates into a separate cluster. Additionally, the presence of a single *Xci* isolate, as well as some *Xcc* isolates that did not group with any specific cluster, motivated the inclusion of the Portuguese set of isolates into a more worldwide population of *X. campestris*. Upon the construction of the NJ tree using the data from the 75 isolates, the overall topology was retained, indicating that the Portuguese isolates are evenly distributed among the population in spite of the high genetic diversity that characterizes them (Cruz *et al.*, 2017).

In addition to the two previously described genetic groups within *Xcc* (Fargier *et al.*, 2011), evidence of high intrapathovar diversity was also found with the establishment of a third genetic group within this pathovar. However, considering its position in the phylogenetic NJ tree, this set of isolates (Xcc3) appeared more closely related to *Xci* and *Xcr* than to the remaining *Xcc* clusters. Two additional clusters, one comprising *Xci* isolates and another grouping *Xcr* isolates, were also disclosed.

The unexpected grouping of some isolates from different pathovars in the same clusters of the NJ tree motivated the application of other phylogenetic analyses. Unlike phylogenetic trees, phylogenetic networks allow the representation of conflicting signals or alternative phylogenetic histories (Bryant & Moulton, 2004). Bifurcating tree-like representations are often misinterpreted as depicting the precise evolutionary history of the isolates, easily leading to false conclusions due to

possible effects of recombination (Hall & Barlow, 2006). Graph-based approaches, such as neighbour-nets, should be preferred when recombination is high, since they scale well and provide a rapid overview of the structure, while pinpointing the origin of the complexity. The intricacy of the neighbour-net constructed for the concatenated dataset from the *X. campestris* isolates (Figure 12B) is indicative of frequent recombination events. Contrastingly, the I_A value significantly different from zero and suggesting linkage disequilibrium indicates that recombination events are rare, a characteristic of clonal populations. However, in several pathogenic bacteria, the ability for localized recombination may be essential to maintain useful variation within the population against the purifying effects of selection (Maynard-Smith *et al.*, 1993). Although frequent recombination mobilizes genetic variation, it may not be sufficient to fully disrupt clonality of the population.

The use of the goeBURST algorithm to construct a MS tree of population structure provided a global overview of the relatedness of the studied isolates and how the defined clusters are connected to each other. Allelic profile based inference of phylogenetic relationships has been proved more reliable than nucleotide based approaches, since replacement of an allele by recombination is scored as a single event (Francisco *et al.*, 2009; Guo *et al.*, 2015).

The addition of Portuguese isolates to the previously studied set of *X. campestris* isolates revealed the presence of new alleles for all loci, many of them represented by a single isolate, and doubled the number of described STs. As expected, no isolates from different pathovars were assigned to same ST, confirming the adequacy of this MLST scheme for the distinction of *X. campestris* pathovars.

A high level of genetic diversity was found among all pathovars. ST5, corresponding to ST17 described by Fargier *et al.* (2011), remained the most prevalent ST and was considered the founder of the single clonal complex described in this study. Although it has no formal definition, a clonal complex comprises genetically related bacteria and constitutes a robust typing unit useful for epidemiological analyses (Urwin & Maiden, 2003). The emergence of clonal complexes is a consequence of a fitness advantage or genetic drift (Francisco *et al.*, 2009): the frequency of a fit genotype increases in the population and it becomes a founder clone; subsequent diversification by mutation and/or recombination leads to the formation of a cluster comprising phylogenetically close strains (Feil *et al.*, 2004). For this reason, founder STs are usually isolated from distinct geographic locations and over the years (Urwin & Maiden, 2003). In the present study, a similar population structure is revealed for *X. campestris*. The founder ST5 comprises isolates that were collected between 1967 and 2002, from diverse locations and hosts. As expected, most of the remaining isolates appear genetically related and its origin can be traced back to the founder ST5.

Although there is evidence of a high degree of genetic diversity among and across each pathovar, there was a clear genetic separation of *X. campestris* pathovars (Figure 13), as suggested by other studies (Fargier *et al.*, 2011). *X. campestris* pathovar classification appears to coincide with a genetic adaptation conferring the ability to colonize different hosts. Host shifts triggering genetic adaptation, tightly associated with positively selected genes, especially those associated with plant infection, are evident in genus *Xanthomonas* (Lu *et al.*, 2008; Huang *et al.*, 2015). The existence of xanthomonads pathogenic to very different host groups appears to have relied on the acquisition of foreign genes conferring ecological fitness, supporting the idea that xanthomonads followed ecological speciation, which describes species arising from ecological diversification (Vos, 2011).

Contrary to evidences for pathovar differentiation in other xanthomonads, like *X. citri* (Huang *et al.*, 2015), host adaptation is not so clear for *X. campestris* pathovars. In this study, the existence of one intermediate *Xcc* ST (ST36) sharing similarities with a *Xcr* ST suggests that these pathovars are more closely related to each other than to *Xci*. This gradient of genetic relatedness seems to be associated to the host range of each pathovar. While *Xci* appears only to be pathogenic on ornamental Brassicaceae, *Xcc* and *Xcr* have a partially overlapping host range: *Xcc* affects mostly Brassicaceous hosts, whereas *Xcr*, while affecting the same hosts, is also pathogenic on Solanaceous hosts (Fargier & Manceau, 2007). These results corroborate recent findings suggesting that instead of causing a host shift, the genetic divergence between *Xcc* and *Xcr* conferred the latter strains the ability to explore additional hosts, providing a broader host range (Huang *et al.*, 2015).

Despite these findings, correlating STs to specific host species or groups was not possible for each pathovar, since isolates appear to be able to infect all members of a given host family, independently of ST (Figure 13a). It is important to notice that for 17 of the 75 isolates studied, varietal information was not available, making it difficult to fully overlap STs and host of origin and establish a clear correlation between genetic and phytopathogenic diversity. Additionally, the four loci analysed, while sufficient for pathovar distinction, may not be informative enough to reveal the underlying host shifts. Screening of type III secretion system components and effectors involved in pathogenesis could be helpful to explain host adaptation within *X. campestris*, as described for other plant pathogenic bacteria, (Grant *et al.*, 2006; Sarkar *et al.*, 2006; Guy *et al.*, 2013).

No geographical correlation between STs was evident (Figure 13b), suggesting that the evolutionary path of *X. campestris* was not shaped by natural barriers. Considering that *X. campestris* is mostly a seed-borne pathogen whose dispersal often follows commercial routes across the globe, it is predictable to find closely related isolates in distant locations, corresponding to a globally spread genetic background (Gitaitis & Walcott, 2007). However, Portuguese isolates were considerably more diverse than isolates from other regions, as indicated by the presence of 26 unique STs. This country is

an important centre of domestication of *B. oleracea* crops and *B. oleracea* var. *costata* (Portuguese cabbage) and *B. oleracea* var. *acephala* (Galega kale), while sharing the same genetic background, form a distinct and homogeneous group, unique within the species (Dias *et al.*, 1992). Centers of origin and areas where wild progenitors cohabit with cultivated plants have long been considered as important sources for disease resistance genes (Leppik, 1970), forcing plant pathogens to adapt, in a process analogous to co-evolution (Rausher, 2001). Thus, these regions may represent natural reservoirs of genetic diversity, the driving force of evolution for both biological systems.

Similarly, no correspondence between STs and *Xcc* races was clear (Figure 13c), suggesting that the molecular basis of this type of host-specificity is not associated with genotype background and/or cannot be revealed by MLST analysis of housekeeping genes. In fact, molecular explanation of race structure has been defined on the basis of the interaction of virulence genes present in the pathogens' genome with the respective resistance genes present on the hosts' genome (Vicente *et al.*, 2001; Fargier & Manceau, 2007; He *et al.*, 2007).

In this chapter, the goeBURST algorithm applied to MLST data proved to be a powerful tool for population structure analysis displaying the molecular epidemiology patterns of this plant pathogen. In fact, the complexity of the population structure of *X. campestris* was revealed, suggesting frequent gene flow both across and within pathovars and supporting the conception that this group of pathogens is highly heterogeneous. The global widespread of the genetic variants is consistent with the epidemiology of seed-borne diseases. *Xcc* races were evenly distributed within the pathovar, suggesting that this type of host-specificity is the result of a complex host-pathogen interaction that goes beyond the information provided by current approaches. Portuguese *X. campestris* isolates represent a significant input of genetic variation and constitute a highly informative and representative sample of *X. campestris* worldwide population.

References

Abdulgader SM, Shittu AO, Nicol MP, Kaba M, 2015. Molecular epidemiology of Methicillin-resistant *Staphylococcus aureus* in Africa: a systematic review. *Frontiers in microbiology* **6**, 1–21.

Bryant D, Moulton V, 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* **21**, 255–265.

Castillo JA, Greenberg JT, 2007. Evolutionary dynamics of *Ralstonia solanacearum*. *Applied and environmental microbiology* **73**, 1225–1238.

Cruz J, Tenreiro R, Cruz L, 2017. Assessment of diversity of *Xanthomonas campestris* pathovars affecting cruciferous plants in Portugal and disclosure of two novel *X. campestris* pv. *campestris* races. *Journal of Plant Pathology* **99**, 403–414.

Dias JS, Lima MB, Song KM *et al.*, 1992. Molecular taxonomy of Portuguese tronchuda cabbage and kale landraces using nuclear RFLPs. *Euphytica* **58**, 221–229.

Doyle JJ, Gaut BS, 2000. Evolution of genes and taxa: a primer. *Plant molecular biology* **42**, 1–23.

Fargier E, Fischer-Le Saux M, Manceau C, 2011. A multilocus sequence analysis of *Xanthomonas campestris* reveals a complex structure within crucifer-attacking pathovars of this species. *Systematic and applied microbiology* **34**, 156–165.

Fargier E, Manceau C, 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**, 805–818.

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG, 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of bacteriology* **186**, 1518–1530.

Francisco AP, Bugalho M, Ramirez M, Carriço J a, 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* **10**, 1–15.

Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA, 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC bioinformatics* **13**, 1–10.

Gitaitis R, Walcott R, 2007. The epidemiology and management of seedborne bacterial diseases. *Annual review of microbiology* **45**, 371–397.

Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL, 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annual review of microbiology* **60**, 425–449.

Guo C, Yang X, Wu Y *et al.*, 2015. MLST-based inference of genetic diversity and population structure of clinical *Klebsiella pneumoniae*, China. *Scientific Reports* **5**, 1–6.

Guy E, Genissel A, Hajri A *et al.*, 2013. Natural genetic variation of *Xanthomonas campestris* pv. *campestris* pathogenicity on *Arabidopsis* revealed by association and reverse genetics. *mBio* **4**, e00538-12.

Hall BG, Barlow M, 2006. Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. *Annals of epidemiology* **16**, 157–169.

Hauben L, Vauterin L, Swings J, Moore ER, 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *International journal of systematic bacteriology* **47**, 328–335.

He Y-Q, Zhang L, Jiang B-L *et al.*, 2007. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome biology* **8**, R218.

Huang C-L, Pu P-H, Huang H-J *et al.*, 2015. Ecological genomics in *Xanthomonas*: the nature of genetic adaptation with homologous recombination and host shifts. *BMC genomics* **16**, 1–13.

Huson DH, Bryant D, 2006. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**, 254–267.

Lelliot RA, Stead D, 1987. *Methods for the diagnosis of bacterial diseases of plants*. Oxford: Blackwell Scientific Publications Ltd.

Leppik EE, 1970. Gene Centers of Plants as Sources of Disease Resistance. *Annual Review of Phytopathology* **8**, 323–344.

Lu H, Patil P, Van Sluys M-A *et al.*, 2008. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PloS one* **3**, e3828.

Maynard-Smith J, Smith NH, O'Rourke M, Spratt BG, 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences* **90**, 4384–4388.

Nei M, Gojobori T, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution* **3**, 418–426.

Parkinson N, Cowie C, Heeney J, Stead D, 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *International Journal of Systematic and Evolutionary Microbiology* **59**, 264–274.

Rausher MD, 2001. Co-evolution and plant resistance to natural enemies. *Nature* **411**, 857–864.

Sarkar SF, Gordon JS, Martin GB, Guttman DS, 2006. Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* **174**, 1041–1056.

Sarkar SF, Guttman DS, 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology* **70**, 1999–2012.

Urwin R, Maiden MCJ, 2003. Multi-locus sequence typing: a tool for global epidemiology. *Trends in microbiology* **11**, 479–487.

Vicente JG, Conway J, Roberts SJ, Taylor JD, 2001. Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathovars. *Phytopathology* **91**, 492–499.

Vicente JG, Everett B, Roberts SJ, 2006. Identification of isolates that cause a leaf spot disease of Brassicas as *Xanthomonas campestris* pv. *raphani* and pathogenic and genetic comparison with related pathovars. *Phytopathology* **96**, 735–745.

Vos M, 2011. A species concept for bacteria based on adaptive divergence. *Trends in microbiology* **19**, 1–7.

Yuan X, Morano L, Bromley R, Spring-Pearson S, Stouthamer R, Nunney L, 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. *Phytopathology* **100**, 601–611.

III. *In vivo* transcriptome profiling of *Xanthomonas campestris* pv. *campestris*

THIS PART IS DIVIDED INTO THE FOLLOWING CHAPTERS:

THE RESULTS DESCRIBED IN THIS PART WERE INCLUDED IN THE FOLLOWING COMMUNICATIONS:

- **Cruz, J.**, Amaral, A. J., Gama-Carvalho, M., Tenreiro, R., Cruz, L. (2017) *In vivo* transcriptome profiling of *Xanthomonas campestris* pv. *campestris* strains with contrasting virulence. *Deciphering plant-microbe interactions for sustainable crop production.* Workshop of the Effectome and Resistance Networks (Banyuls sur Mer, France).
- **Cruz, J.**, Amaral, A. J., Gama-Carvalho, M., Tenreiro, R., Cruz, L. (2017) Profiling the in vivo transcriptome of *Xanthomonas campestris* pv. *campestris* strains with contrasting virulence. *MicroBiotec'17 – Portuguese Congress of Microbiology and Biotechnology* (Porto, Portugal).

1 Virulence assessment of Portuguese *Xanthomonas campestris* pv. *campestris* isolates

1.1 Introduction

Throughout the history of plant pathology, both qualitative and quantitative aspects of plant disease have been analyzed. Although terminology has been an issue for the past decades, pathogenicity can be described as the qualitative ability to cause disease to a given host, while virulence corresponds to the quantifiable damage caused to the infected host (Sacristán & García-Arenal, 2008; Surico, 2013).

Both virulence and pathogenicity have ecological consequences, affecting disease dynamics and epidemiology, as well as shaping pathogen and host community structure (Barrett *et al.*, 2009). From an epidemiological stand-point, pathogenicity plays a key role in the adaptation of pathogens to new hosts, especially in the case of emerging pathogens and/or in the face of environmental changes (Anderson *et al.*, 2004), whereas virulence modulates pathogen fitness, thus shaping its evolution and further influencing pathogen interaction with potential hosts (Pariaud *et al.*, 2009). Evolution of pathogenicity and virulence are not detachable and both features are important driving forces behind co-evolution of host-pathogen systems (Sacristán & García-Arenal, 2008), although evolutionary patterns differ between specialized and generalist pathogens (Lannou, 2012). For more specialized pathogens, displaying a restricted host range, natural selection will favor high levels of virulence whenever it leads to increased growth and among-host transmission (Lannou, 2012). On the other hand, the broader host range of generalist and opportunistic pathogens rewards them with more opportunities for transmission and persistence, thus associating decreased pathogenicity with increased virulence (Lannou, 2012).

Plant pathologists have focused mainly on pathogenicity molecular mechanisms, mostly due to the molecular determinism presented by Flors' gene-for-gene model for compatible or incompatible reactions. Following this model, the existence of direct recognition of a pathogens' avirulence gene (*avr*) by a corresponding host resistance gene (R) leads to qualitative resistance. However, qualitative resistance introduced by the artificial deployment of R genes is rapidly rendered ineffective in the widespread modern monocultural cropping systems, mostly due to its narrow genetic base, which results in a strong directional selection on pathogens (Zhan *et al.*, 2015).

On the other hand, virulence is often assumed as a polygenic trait and therefore quantitative adaptation to the host is theoretically expected to be slower than the acquisition of additional qualitative virulence factors, and quantitative plant resistances are generally expected to be more durable than qualitative resistances (Pariaud *et al.*, 2009). For these reasons, there is an increasing

interest in quantitative resistance and in the combination of quantitative and qualitative resistance strategies, although experimental work on virulence and its contribution for quantitative resistance is somewhat embryonic (Lannou, 2012; Zhan *et al.*, 2015).

Virulence has no apparent benefit for pathogens and is thus often regarded as an inevitable consequence of within-host multiplication (Lenski & May, 1994). Following this trade-off hypothesis, a positive correlation between virulence and within-host multiplication and transmission is thus expected. However, higher virulence would ultimately lead to greater host mortality or lack of fitness, negatively correlated with between-host transmission. One of the pitfalls of testing this hypothesis is the difficulty in the assessment of virulence. While for animal pathogens virulence (or infectivity) can be a measure of host mortality, for plant pathogens, the effects of virulence may not be immediate (Sacristán & García-Arenal, 2008).

Measuring virulence in plant pathogens rarely integrates all individual quantitative traits of the pathogens life cycle and usually a combination of individual traits is undertaken. Given host-pathogen interaction characteristics, virulence can be measured as infection efficiency (number of lesions), sporulation rate, infectious period, latent period, lesion size or even toxin production (Pariaud *et al.*, 2009). For bacterial plant diseases, lesion number or lesion size after artificial infection are the most measured attributes, when assessing the virulence of a parenchymatous or vascular infection, respectively. Although some advances have been made in the field of machine vision, with the incorporation of chlorophyll fluorescence imaging, hyperspectral imaging or thermal imaging, image-based phenotyping of plant disease still mostly relies on visible light imaging (Mutka & Bart, 2015).

Particularly severe in warm, humid regions, black rot disease, affecting all cultivated Brassicas, is one of the most important bacterial diseases worldwide (Mansfield *et al.*, 2012; Vicente & Holub, 2013). The gene-for-gene model applied to *Xanthomonas campestris* pv. *campestris* (*Xcc*), the causative agent of black rot disease, has led to the establishment of a race structure, based on the interaction of *avr*-R gene pairs (Vicente *et al.*, 2001; Fargier & Manceau, 2007). Despite the characterization of resistance in several genotypes of *Brassica* spp., no resistance genes have been successfully cloned (Vicente & Holub, 2013; Lee *et al.*, 2015). Quantitative virulence traits leading to the development of quantitative resistance in susceptible hosts has also been largely unexplored in this host-pathogen interaction.

In this chapter, the differential virulence assessment of a set of Portuguese *Xcc* isolates was carried out, with the main goals of i) assessing virulence of *Xcc* against selected hosts by measuring the infected leaf area and ii) describing the relationships between pathogenicity and virulence in this widely-distributed plant pathogenic bacterium.

1.2 Materials and methods

1.2.1 Bacterial isolates and virulence assessment

To determine the virulence of Portuguese *Xcc* isolates, a set of 26 *Xcc* isolates was inoculated on Savoy cabbage (*B. oleracea* var. *sabauda* cv. 'Wirosa F1' – Sommers Seeds, BE), Portuguese cabbage (*B. oleracea* var. *costata* cv. 'Beira F1' – Bejo Zaden, NL) and Galega kale (*B. oleracea* convar. *Acephala* var. *sabellica* cv. 'Bonanza F1' – Bejo Zaden, NL). For each isolate/host combination, inoculations were performed in triplicate, by spraying a bacterial water suspension (10^8 cfu.ml^{-1}; OD_{600}=0.1) directly onto the leaf surfaces, until run-off, as described in Chapter 2.1. After inoculation, plantlets were maintained in closed polyethylene bags for 48h to promote the infection process and kept for 15 days at 24/18°C with a 16h-light/8h-dark photoperiod (RH>80%) and checked daily for symptom development. At the end of the assay, a total of eight leaves – five older leaves and three younger leaves per plant – were randomly collected and photographed (Figure 16D). For each leaf, the total and the diseased visible area were digitally measured and recorded, using the measurement tool of Adobe Photoshop CS5 software (Adobe Systems, San Jose, United States of America), and the percentage of infected leaf area (%ILA) was calculated. In agreement with the conclusions of the exploratory data analysis described below, virulence assessment for the set of 26 *X. campestris* pv. *campestris* (*Xcc*) isolates collected in Portugal was carried out based on total infected leaf area (%ILA) for each of the three hosts.

1.2.2 Data analysis

Exploratory data analysis using the percentage of infected leaf area (%ILA) was carried out, using data for total %ILA, young %ILA (measured on three young leaves) and old %ILA (measured on five old leaves) for each host.

Pearson's correlation coefficient was used to assess the correlations between those variables computing and correlation matrix visualization was carried out using R package **ggcorrplot** (https://cran.r-project.org/package=ggcorrplot). Principal Component Analysis (PCA) was computed using the **stats** built-in R package, while extraction and visualization of PCA results was carried out using R packages **factoextra** (https://cran.r-project.org/package=factoextra) and **scatterplot3d** (https://cran.r-project.org/package=scatterplot3d).

1.3 Results and discussion

1.3.1 Exploratory data analysis

The average percentage of infected leaf area (%ILA) caused by each isolate upon inoculation on the hosts tested is recorded in Table 11 (the individual measurements are in Appendix).

Table 11 – Portuguese *Xanthomonas campestris* pv. *campestris* isolates used in this study, their race identification and host of origin. Virulence, measured as the mean percentage of infected leaf area (%ILA) in young leaves (%ILA YL), old leaves (%ILA OL) and total (%ILA Total) is also given for tested hosts – *B. oleracea* var. *sabauda* cv. 'Wirosa F1', *B. oleracea* convar. *acephala* var. *sabellica* cv. 'Bonanza F1' and *B. oleracea* var. *costata* cv. 'Beira F1'. Bold – variation coeficient ≤30%; Italic – variation coeficient ≥ 60%. np – not pathogenic.

		%ILA Beira F1			%ILA Bonanza F1			%ILA Wirosa F1		
CPBF #	**Host of origin**	**YL**	**OL**	**Total**	**YL**	**OL**	**Total**	**YL**	**OL**	**Total**
RACE 4										
46	*B. oleracea* convar. *acephala* var. *sabellica*	4.09	*38.71*	*17.07*	*5.67*	4.12	*5.09*	*5.36*	15.76	*9.26*
RACE 6										
208	*B. oleracea* var. *costata*	np	np	np	*9.62*	*16.01*	*12.02*	6.34	**100.00**	*41.46*
213	*B. oleracea* var. *costata*	0.00	*4.76*	*1.78*	7.03	*40.30*	*19.51*	*3.32*	*2.32*	*2.95*
278	*B. oleracea*	*17.78*	**100.00**	*48.61*	*13.09*	*73.76*	*35.84*	6.15	**100.00**	*41.35*
279	*B. oleracea*	*5.17*	80.70	*33.49*	*6.38*	*56.40*	*25.13*	*19.48*	*44.47*	*28.85*
589	*B. oleracea* convar. *botrytis* var. *botrytis*	15.31	23.97	18.56	16.16	*54.87*	*30.68*	**5.40**	78.32	*32.75*
667	*B. oleracea*	21.48	*47.33*	*31.18*	11.67	*37.24*	*21.26*	*8.34*	31.44	*17.00*
1126	*B. oleracea*	*22.57*	37.80	*28.28*	*3.84*	24.00	*11.40*	*16.70*	*71.71*	*37.33*
1135	*B. oleracea* convar. *capitata* var. *alba*	*19.43*	**32.34**	24.27	*4.24*	19.40	*9.93*	12.95	29.27	*19.07*
1136	*B. oleracea* convar. *capitata* var. *alba*	*18.00*	36.91	*25.09*	10.27	**5.60**	8.52	8.15	*46.92*	*22.69*
1175	*B. oleracea* convar. *botrytis* var. *botrytis*	*25.64*	*48.31*	*34.14*	*13.64*	7.72	*11.42*	5.80	10.22	7.46
1176	*B. oleracea* convar. *botrytis* var. *botrytis*	19.01	*32.80*	*24.19*	*6.60*	**11.87**	8.58	*10.95*	**11.17**	11.03
RACE 7										
140	*B. oleracea* var. *costata*	*8.19*	**100.00**	*42.62*	np	np	np	**20.39**	77.32	*41.74*
210	*B. oleracea* convar. *botrytis* var. *botrytis*	*9.21*	80.79	*36.05*	np	np	np	18.86	*47.05*	*29.43*
211	*B. oleracea* convar. *botrytis* var. *botrytis*	np	np	np	np	np	np	21.96	**22.96**	22.34
212	*B. oleracea* var. *costata*	np	np	np	np	np	np	7.41	14.72	*10.15*
216	B. *oleracea* convar. *botrytis* var. *botrytis*	*1.18*	*45.10*	*17.65*	np	np	np	*18.30*	80.15	*41.49*
329	*B. oleracea* var. *costata*	18.87	80.51	*41.99*	np	np	np	*12.52*	*12.19*	*12.40*
332	*B. oleracea* var. *costata*	*22.18*	*56.94*	*35.21*	*7.96*	*73.75*	*32.63*	*10.65*	18.80	*13.71*
668	*B. oleracea*	np	np	np	np	np	np	np	np	np
824	*B. oleracea*	18.55	*14.58*	*17.07*	5.69	17.07	*9.96*	*11.17*	**26.94**	*17.09*
RACE 10										
330	*B. oleracea* var. *costata*	*7.07*	37.81	*18.60*	*5.46*	*73.73*	*31.06*	*10.90*	*38.45*	*21.23*
331	*B. oleracea* var. *costata*	4.94	**100.00**	*40.59*	np	np	np	*10.03*	**100.00**	*43.77*
604	*B. oleracea* convar. *capitata* var. *sabauda*	*11.65*	**23.00**	*15.91*	13.91	**100.00**	*46.20*	*5.08*	**100.00**	*40.68*
RACE 11										
489	*B. oleracea* var. *costata*	np	np	np	np	np	np	np	np	np
1198	*B. kaber* var. *pinnatifida*	np	np	np	*2.93*	3.74	*3.23*	np	np	np

Presence of symptoms in younger leaves born after inoculation time can be indicative of a higher tolerance to *Xcc*, since the pathogen did not cause heavy infection damages permitting vasculature colonization. For cv. 'Bonanza', correlation between %ILA on young leaves and old leaves was higher than for the remaining hosts, suggesting that on this host the colonization of the vasculature occurs faster (Figure 14). Interestingly, this host recorded the lowest average %ILA, suggesting higher resistance to *Xcc*. It is important to notice that there is a higher correlation between old %ILA and total %ILA than between young %ILA and total %ILA for all three hosts. As expected, this finding indicates that, although there is a vascular transmission of the pathogen to the younger leaves,

the %ILA on the older leaves is a good measurement of global virulence. Therefore, in future analysis, %ILA measurements on the tested hosts will be carried out using only the existing leaves at the time of the inoculation. Additionally, no significant correlation was found between hosts, neither for total %ILA or for young or old leaves %ILA, indicating that each *Xcc* tested isolate affects this set of hosts in an unrelated manner.

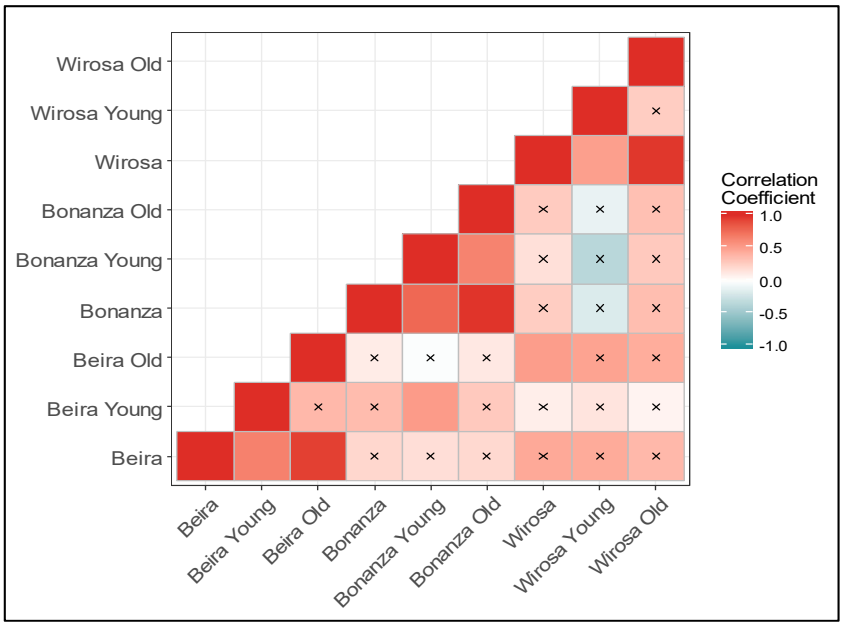


Figure 14 – Pearson's correlation matrix, highlighting the correlations between the total Percentage of Infected Leaf Area (%ILA) for each tested host (Beira, Bonanza and Wirosa), %ILA for young leaves and %ILA for old leaves. (x) not significant (*p*-value > 0.05).

After principal component analysis (PCA) was performed, dimensions 1-3 explained 42.1%, 27.8% and 15.8% of the total data variance, respectively, resulting in a cumulative variance of 85.7%.

The correlation between the original variables and the dimensions retained after PCA are depicted in Figure 15A. While dimension 1 represents a growing continuum of virulence, dimension 2 is able to inversely correlate the isolates that affect Beira and Wirosa, with those that only affect Bonanza. Additionally, dimension 3 can further separate the isolates that affecting both Beira and Wirosa, cause more symptoms in Beira than in Wirosa.

The contributions of the original variables to the newly determined dimensions were also assessed and are depicted in Figure 15B. Original variables with a contribution above the expected, i.e. if all variables were equally important, are considered important for the dimension in question. Although contributions vary for each new dimension, it is evident that the variables corresponding to the total ILA per host (Wirosa, Beira and Bonanza) have the higher contribution for the new dimensions created, thus reinforcing that distinct measurements of %ILA on young versus old leaves are unnecessary to explain the variance in the original data, since there is no clear age-related resistance.

Age-related resistance to pathogens, as shown by different levels of resistance depending on the age of the plant or of the plant tissues, has been described for many plants, and is often positively correlated with plant maturity (Panter & Jones, 2002). For *Arabidopsis thaliana* infected with *Pseudomonas syringae* pv. *tomato* and *P. syringae* pv. *maculicola*, resistance in mature plants was triggered by the vegetative-floral phase transition and was accompanied by a reduced bacterial growth and absence of symptoms (Kus *et al.*, 2002). Similarly, for *Xanthomonas oryzae* pv. *oryzae* affecting rice plants containing the resistance gene Xa21, age-related resistance was also reported, as a reduction in the length of disease lesions with increasing plant age (Mazzola *et al.*, 1994). For the pathosystems analyzed during this work, the results obtained do not suggest the existence of age-related resistance phenomena. However, in order to assess the existence of age-related resistance in the studied pathosystems, inoculations should be performed in plants with different ages or in differing phenological states.

Although the new dimensions achieved using PCA confirm the correlations between variables obtained using Pearson's correlation, they did not highlight any clustering pattern of the isolates, based on additional features, such as race or host of origin as depicted in the three-dimensional projection of the isolates on the principal components in Figure 15C and D.

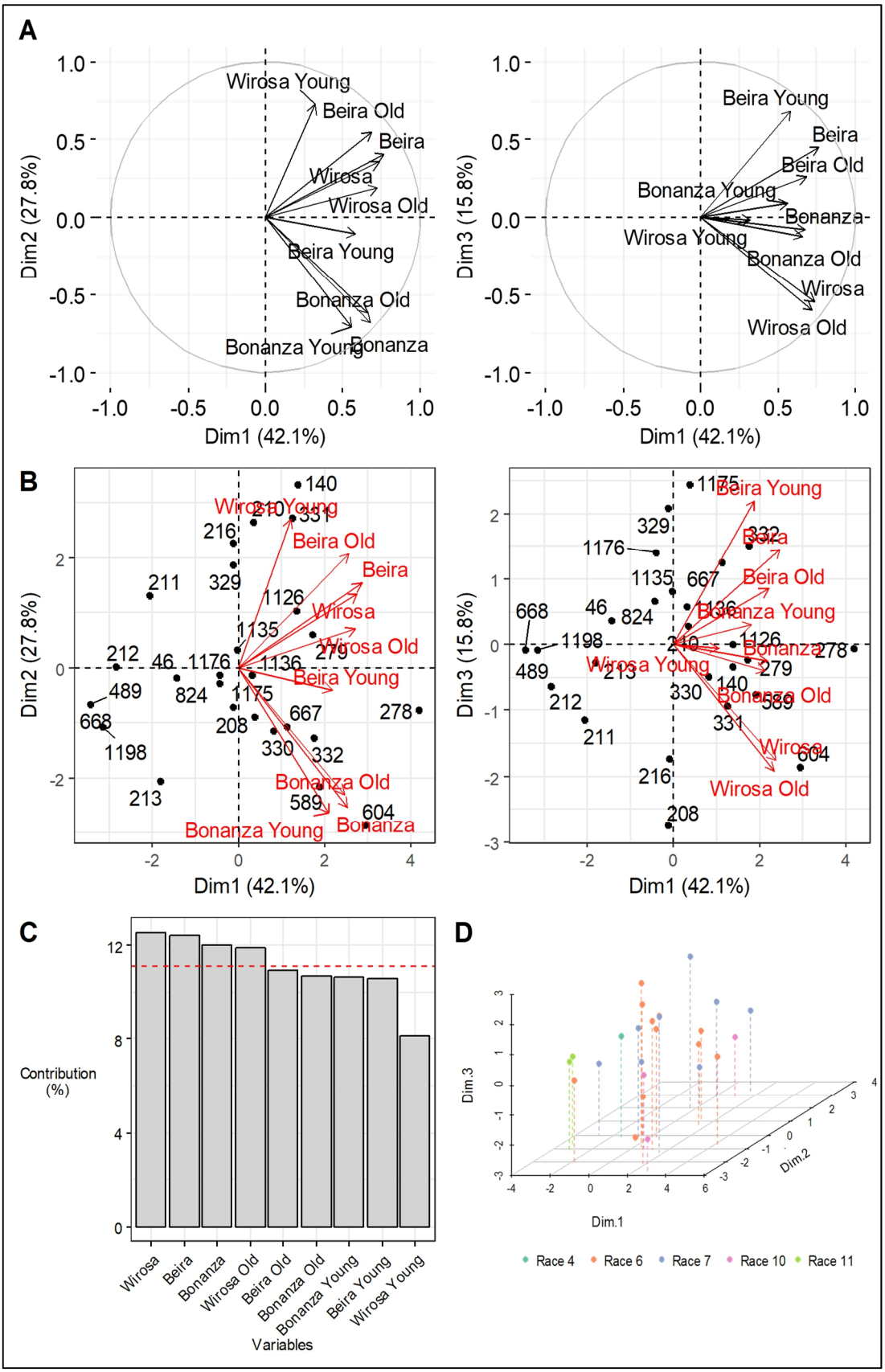


Figure 15 – Principal Component Analysis (PCA) of the percentage of infected leaf area. (A) Correlation circles of the original variables with the retained PCA dimensions. (B) Biplots of original variables and individuals on the new dimensions created after PCA. (C) Contribution of original variables to PCA dimensions 1, 2 and 3. Red dashed line indicates the expected contribution. (D) Three-dimensional plot of *Xanthomonas campestris* pv. *campestris* (*Xcc*) isolates, on the retained three dimensions after Principal Component Analysis. *Xcc* races are highlighted.

1.3.2 Virulence assessment

Among the 26 Portuguese *Xcc* tested isolates, 15 infected the three selected hosts. Of the remaining 11 isolates, 6 infected two hosts and 3 caused symptoms on a single host. Additionally, CPBF668 and CPBF489, belonging to race 7 and 11, respectively, did not cause any symptoms on the tested hosts, although they were previously proved pathogenic on other Brassicaceae hosts. The isolate races did not correlate to the number of hosts they infected nor to their virulence in any of the hosts tested, suggesting that virulence is an intrinsic feature of each isolate, independent of pathogenicity reflected by the host range.

The results of this assessment are in accordance to what has been described for other bacterial plant pathogens, such as *Pseudomonas viridiflava*, affecting the model plant *Arabidopsis thaliana* (Goss & Bergelson, 2006). In that study, virulence assessment by recording differential symptom evolution (diseased leaf area, scored by eye) in a cross inoculation of 23 bacterial isolates in 35 plant lines, revealed significant variation for different pathogen clades (Goss & Bergelson, 2006).

As depicted in Figure 16, of the isolates causing black rot disease symptoms in all tested hosts, CPBF278 was found to be the most virulent isolate affecting 42% of total leaf area (H-vir isolate) while CPBF213 was considered the least virulent affecting only 8% of total leaf area (L-vir isolate). Since both these isolates belong to race 6 and could infect the three additional hosts tested, it is possible to conclude they have the same host range. Despite the high variability of the %ILA individual measurements, the 95% confidence intervals calculated for each isolate do not overlap, thus suggesting that the estimated differences are significant (Figure 16B). Therefore, these isolates were selected as the lowest and highest virulence strains, and can thus be used as the contrasting extremes of the virulence spectrum of equally pathogenic *Xcc* strains.

In terms of host susceptibility, for all isolates tested, the most resistant host was cv. 'Bonanza' with an average ILA of 12%, while cv. 'Wirosa' and cv. 'Beira' were similarly susceptible, showing an average ILA of 21.7% and 21.2%, respectively. These findings suggest that for virulence determination purposes, cv. 'Wirosa' and cv. 'Beira' can be alternatively used.

Overall, these results reinforce the perception that the current concept of pathogenicity does not reflect the actual quantifiable damage that an isolate can cause to susceptible hosts, and therefore should not be used to such purpose *per se*. Additionally, the co-evolution between pathogens and their hosts plants is usually in favor of the pathogens, due to their shorter generation times, associated to a restricted host genetic background resulting from the modern cropping systems (Zhan *et al.*, 2015). The development of sustainable control strategies and measures must therefore minimize the pathogen evolutionary potential, shifting it to the host plants. In this context, and considering its

shaping effect on pathogen fitness (Sacristán & García-Arenal, 2008), understanding virulence in plant pathogens must also be a key step for the development of control measures.

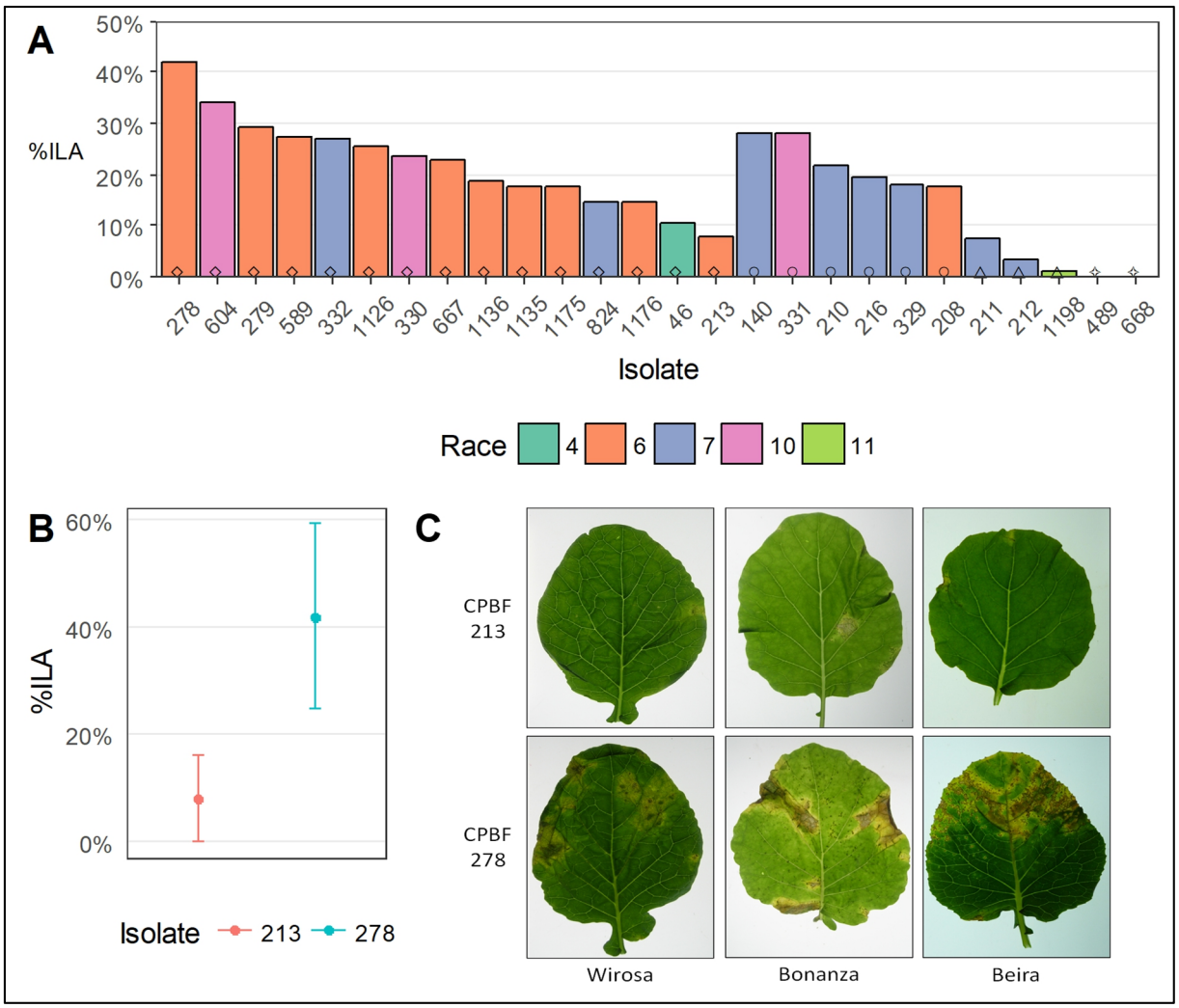


Figure 16 – Virulence assessment of the Portuguese *Xanthomonas campestris* pv. *campestris* isolates against hosts tested. (A) Percentage of Infected Leaf Area (%ILA). ◇ Isolates infecting three hosts; ○ Isolates infecting two hosts; △ Isolates infecting one host; ✧ Non-pathogenic isolates. (B) Confidence intervals for isolates CPBF213 and CPBF278. (C) Symptoms caused by *Xanthomonas campestris* pv. *campestris* isolates CPBF213 and CPBF278 inoculated on *Brassica oleracea* var. *sabauda* cv. 'Wirosa', *B. oleracea* convar. *acephala* var. *sabellica* cv. 'Bonanza' and *B. oleracea* var. *costata* cv. 'Beira'.

In this study, the tests carried out using a set of Portuguese *Xanthomonas campestris* isolates revealed that virulence within *X. campestris* pv. *campestris* (*Xcc*) races was not homogeneous. Furthermore, *Xcc* isolates belonging to the races with a broader host range – races 6 and 7 – did not necessary correspond to the ones with higher virulence. The reliable determination of the virulence of each *Xcc* isolate therefore appears of utmost importance under field conditions, since virulence patterns may play a more relevant role in disease progression than the actual host range determined by race. In addition to host range, virulence of the strains should also be measured as a component of pest risk analysis.

References

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P, 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in ecology & evolution* **19**, 535–544.

Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J, 2009. Continua of specificity and virulence in plant host-pathogen interactions: Causes and consequences. *New Phytologist* **183**, 513–529.

Fargier E, Manceau C, 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**, 805–818.

Goss EM, Bergelson J, 2006. Variation in resistance and virulence in the interaction between *Arabidopsis thaliana* and a bacterial pathogen. *Evolution; international journal of organic evolution* **60**, 1562–1573.

Kus J V, Zaton K, Sarkar R, Cameron RK, 2002. Age-related resistance in Arabidopsis is a developmentally regulated defense response to Pseudomonas syringae. *The Plant cell* **14**, 479–490.

Lannou C, 2012. Variation and selection of quantitative traits in plant pathogens. *Annual Review of Phytopathology* **50**, 319–338.

Lee J, Izzah NK, Jayakodi M *et al.*, 2015. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. *BMC Plant Biology* **15**, 1–11.

Lenski RE, May RM, 1994. The Evolution of Virulence in Parasites and Pathogens: Reconciliation Between Two Competing Hypotheses. *Journal of Theoretical Biology* **169**, 253–265.

Mansfield J, Genin S, Magori S *et al.*, 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology* **13**, 614–629.

Mazzola M, Lea J, Nelson R, White F, 1994. Analysis of the Interaction Between *Xanthomonas oryzae* pv. *oryzae* and the rice Cultivars IT24 and IRBB21. *Molecular Plant Pathology* **84**, 392–397.

Mutka AM, Bart RS, 2015. Image-based phenotyping of plant disease symptoms. *Frontiers in Plant Science* **5**, 1–8.

Panter S., Jones D., 2002. Age-related resistance to plant pathogens. *Advances in Botanical Research* **38**, 251–280.

Pariaud B, Ravigné V, Halkett F, Goyeau H, Carlier J, Lannou C, 2009. Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology* **58**, 409–424.

Sacristán S, García-Arenal F, 2008. The evolution of virulence and pathogenicity in plant pathogen populations. *Molecular Plant Pathology* **9**, 369–384.

Surico G, 2013. The concepts of plant pathogenicity, virulence/avirulence and effector proteins by a teacher of plant pathology. *Phytopathologia Mediterranea* **52**, 399–417.

Vicente JG, Conway J, Roberts SJ, Taylor JD, 2001. Identification and origin of *Xanthomonas campestris* pv. *campestris* races and related pathovars. *Phytopathology* **91**, 492–499.

Vicente JG, Holub EB, 2013. *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular plant pathology* **14**, 2–18.

Zhan J, Thrall PH, Papaïx J, Xie L, Burdon JJ, 2015. Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annual Review of Phytopathology* **53**, 19–43.

2 *In vivo* transcriptome profiling of *Xanthomonas campestris* pv. *campestris* strains with contrasting virulence

2.1 Introduction

As for several bacterial plant diseases, the molecular plant-microbe interaction behind black rot disease has been explained by the Flor's gene-for-gene model (Flor, 1956). Initially proposed in the 1950's to explain the host range of flax rust disease, this model has also been used to explain the existence of a race structure within *Xanthomonas campestris* pv. *campestris* (*Xcc*) (Vicente *et al.*, 2002; Fargier & Manceau, 2007). The gene-for-gene model states that disease resistance is observed when the product of any particular resistance gene (R gene) specifically recognizes the product of a specific pathogen avirulence gene (*avr* gene) (Flor, 1956). This model has driven research towards the identification and characterization of genes having a direct effect in pathogenicity, in the hope to find a corresponding R gene leading to plant resistance. On the other hand, plant genomic data have also been critical for the identification of plant cell receptors and their cognate molecules (Bent & Mackey, 2007).

Molecular plant-microbe interaction has been shown to be a conjoint process comprising defense and counter-defense mechanisms. Molecular receptors in the membrane of host cells can recognize universal bacterial determinants, activating the first line of plant immunity (PAMP - triggered immunity) (Jones & Dangl, 2006). In turn, bacteria retort by deploying effector molecules that manipulate the hosts' primary defenses, causing effector-triggered susceptibility (ETS) (Jones & Dangl, 2006; Zhou & Chai, 2008). Eventually, host cells develop additional strategies, recognizing and blocking the neutralizing action of bacterial effectors, in a process of effector-triggered immunity (ETI). This co-evolutionary competition is driven by the constant diversification of bacterial effectors and corresponding plant resistance proteins and affects both natural plant resistance as well as artificially introduced resistance traits (Rausher, 2001).

Genomic approaches to *Xanthomonas* spp. pathogenicity revealed a diverse array of effectors that are secreted and make their way into the host cell, mostly through Type III Secretion Systems (T3SS). Following a comprehensive study using 132 strains of several *X. axonopodis* pathovars, a repertoire-per-repertoire hypothesis has been proposed to explain host specificity and even tissue specificity in the genus, based on the array of Type III Effectors (T3Es) encoded by the bacterial genome (Lemaire *et al.*, 2009). According to this hypothesis, T3E repertoires could explain a pathological convergence of phylogenetically distant strains and would, in its entirety greatly determine the host range (Lemaire *et al.*, 2009). This hypothesis could ultimately lead to the existence of a core effectome

shared by all pathovars of a given *Xanthomonas* species and a variable effectome which would in turn determine host range, differentiating pathovars.

While in *X. axonopodis* the core effectome was found to include 8 genes, comparative genomics using 13 fully sequenced *X. campestris* strains established a rather limited core effectome (Roux *et al.*, 2015). A combination of genomics and transcriptomics data showed that while varying between 13 and 31 predicted T3Es, genomes of *X. campestris* pathovars share only three effectors – XopA, HrpW and HpaA (Roux *et al.*, 2015). The diversity of T3Es among these closely related strains can also help to explain the different tissue specificity when comparing the vascular pathovars *Xcc* and *X. campestris* pv. *incanae* (*Xci*) with the mesophilic *X. campestris* pv. *raphani* (*Xcr*). For example, AvrBs2, XopAL2, XopAM, XopN, XopX21, XopX2, XopZ1 are absent from *Xcr* genomes, being exclusively present in *Xcc* and *Xci* strains (Roux *et al.*, 2015). However, this set of effectors is not present in every strain of *Xcc* and *Xci* analyzed, suggesting that the concept of core effectome within pathovars cannot be fully employed.

Since T3SS is essential for virulence in *Xcc*, it would be expected for T3Es to play a key role in determining virulence. However, individual *Xcc* T3Es have been shown to have a limited contribution to pathogenicity, suggesting the existence of a complex network behind this phenomenon and the redundancy of some effectors. This implies that the T3E repertoire alone cannot completely explain the broad host range and tissue specificity that characterize *X. campestris*. For *Xcc*, the further intrapathovar differentiation into races deepens the plot, since strains belonging to different races were shown to share the majority of T3E candidates (Roux *et al.*, 2015). Additionally, this type of genomic survey of T3E repertoires seems insufficient to explain the fine regulation mechanisms behind different virulence for strains belonging to the same race.

The pitfalls of genome-based approaches used to characterize xanthomonads and other plant pathogenic bacteria showing a broad host range have led the research towards transcriptomics (Kimbrel *et al.*, 2011). In fact, genomic approaches do not consider that host pathogen interaction actively relies on regulation of gene expression in the face of changing environments, contributing for both rapid response mechanisms and evolutionary adaptation. Although transcriptome profiling using microarray technology has been extensively used for plant pathogenic bacteria, next generation sequencing (NGS) approaches have gained momentum only in the last decade, allowing the identification of genes differentially expressed by the same strain under different conditions or by different strains under the same conditions (Liu *et al.*, 2013; Ailloud *et al.*, 2016).

In 2013, Liu *et al.* published a transcriptome profiling of *Xcc* strain 8004, by comparing gene expression upon growth in two distinct culture media – a minimal medium (MMX), typically used to induce T3SS expression, and a rich medium (NYG), used for morphology and biology studies, as well as

for conservation. In addition to the expected differential expression of genes coding for T3SS components and T3Es, several other genes were differentially expressed in MMX, including those involved in basic cellular processes such as nucleotide metabolism or amino acid biosynthesis, suggesting that pathogenicity is a complex process engaging several cellular components (Liu *et al.*, 2013).

Transcriptional reprogramming of *Xcc* has been shown as a response to *Brassica oleracea* xylem sap (de Bernonville *et al.*, 2014). Using cDNA microarrays *Xcc* strain 8004 grown in xylem sap underwent an up-regulation of genes encoding TonB-dependent transporters (TBDTs), as well as of genes important for twitching motility and adhesion, when compared to minimal or rich media (de Bernonville *et al.*, 2014).

Despite these studies, information concerning the actual *in planta* transcriptome of *Xcc* is still largely unexplored. *In vivo* transcriptomic approaches have been successfully used in other plant pathogenic bacteria to characterize pathogenicity in order to identify the molecular basis of host range. In 2016, the transcriptome *Ralstonia solanacearum* strains causing disease to various host plants was determined using an *in planta* RNA-Seq based approach (Ailloud *et al.*, 2016). During this study, the authors showed that gene expression was convergent between the two pathotypes in similar environmental conditions, but the expression of virulence-associated genes diverged during infection of their unique hosts (Ailloud *et al.*, 2016).

These studies confirm the ideas that bacterial gene expression constitutes a dynamic adaptation mechanism during pathogenesis in a host-dependent manner. Using transcriptomics, pathogenicity may correspond to the expression of different sets of genes in different hosts, while virulence may correspond to the different levels of expression of a given gene set.

In the previous chapter, Portuguese *X. campestris* isolates of the same pathovar and race were shown to display very different virulence levels when inoculated on a set of susceptible hosts. This phenotypical contrast was not supported by genomic or phylogenetic data, suggesting that it might be the result of differential gene expression during pathogenesis. In this chapter, RNA-Seq was used to profile the *in planta* transcriptome of two *Xcc* strains displaying contrasting virulence against selected susceptible hosts. The main goals of this work are i) to determine if virulence is host-dependent; ii) to identify novel virulence-associated genes; and iii) to identify gene expression patterns responsible for contrasting virulence in two closely related strains. This work represents the first *in planta* transcriptome study of two different *Xcc* strains using RNA-Seq technology.

2.2 Materials and methods

2.2.1 High-virulence and Low-virulence pathosystems

For this study, two Portuguese *Xcc* isolates were used, representing the extremes of the virulence spectrum - CPBF278, the highest virulence strain (H-vir) and CPBF213, the lowest virulence strain (L-vir).

This selection, carried out using a holistic characterization strategy that integrated phenotypic, genotypic and phylogenetic data, carried out as described in previous chapter, allowed establishing as the most virulent isolate (H-vir isolate), while was considered the least virulent (L-vir isolate).

Given the susceptibility patterns of the three hosts used to determine *Xcc* virulence, plant hosts used in this study included Savoy cabbage (*B. oleracea* var. *sabauda* cv. 'Wirosa F1' – Sommers Seeds, BE) and Galega kale (*B. oleracea* convar. *Acephala* var. *sabellica* cv. 'Bonanza F1' – Bejo Zaden, NL).

Four different pathosystems were established by inoculating each strain onto each host: i) L-vir x Wirosa; ii) L-vir x Bonanza; iii) H-vir x Wirosa; iv) H-vir x Bonanza.

Inoculations of hosts were performed, by spraying a bacterial suspension (10^8 cfu.ml^{-1}; OD_{600}=0.1) directly onto the leaf surface of plantlets with 2-3 true leaves, until run-off. After inoculation, plantlets were maintained in closed polyethylene bags for 48h to promote the infection process. Plants were kept for 15 days at 24/18°C with a 16h-light/8h-dark photoperiod (RH>80%) and checked daily for symptom development. Fourteen days after inoculation, the midribs and petioles of affected leaves were collected, crushed with liquid nitrogen and stored at -80°C until future usage.

2.2.2 RNA isolation and construction of libraries for transcriptome analysis

Total RNA was isolated from the infected plant tissue obtained in 1.2.1 using Direct-zol RNA MiniPrep (Zymo Research). To eliminate genomic DNA contaminations, RNA samples were treated with Amplification Grade DNase I (Invitrogen) at a concentration of 1 U/µg of total RNA.

RNA purity was preliminary assessed using Nanodrop2000, and samples with A260/280 and A260/230 < 1.8 were discarded.

RNA quantity was assessed using a Qubit Fluorometer (Life Technologies) and RNA quality was assessed using an Agilent Bioanalyzer (Agilent Technologies) before proceeding to RNA-Seq. Only the samples with RNA integrity number (RIN) > 7.0 were used.

To guarantee the "functionality" of the isolated RNA, a reverse transcription PCR of the samples was performed. Oligo(dT)-primed cDNA synthesis was carried out using SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen). This cDNA was then used for PCR amplification of a

663bp fragment of the 18s rRNA gene using the primer set NS5F/NS8R (AACTTAAAGGAATTGACGGAAG /TCCGCAGGTTCACCTACGGA) (Pastrik, 2000; Pastrik *et al.*, 2002).

After RNA quantity, quality, integrity and functionality were guaranteed, the samples were sent to the Genomics Core Facility of the European Molecular Biology Laboratory (EMBL - Heidelberg, Germany) where sequencing libraries were prepared and sequenced. Preparation of libraries included rRNA depletion using a mixture of bacterial and plant probes, RNA fragmentation and cDNA synthesis using the Illumina Stranded Kit. The constructed sequencing libraries were sequenced using Illumina Hiseq 2000 platform, generating 75bp single-end reads. Transcriptomes of two independent samples from each of the four established pathosystems were obtained by RNA-Seq, in a total of 8 transcriptomes.

2.2.3 Analysis of RNA-Seq data and differential expression analysis

The workflow of RNA-Seq data analysis is described in Figure 17 and is further detailed bellow.

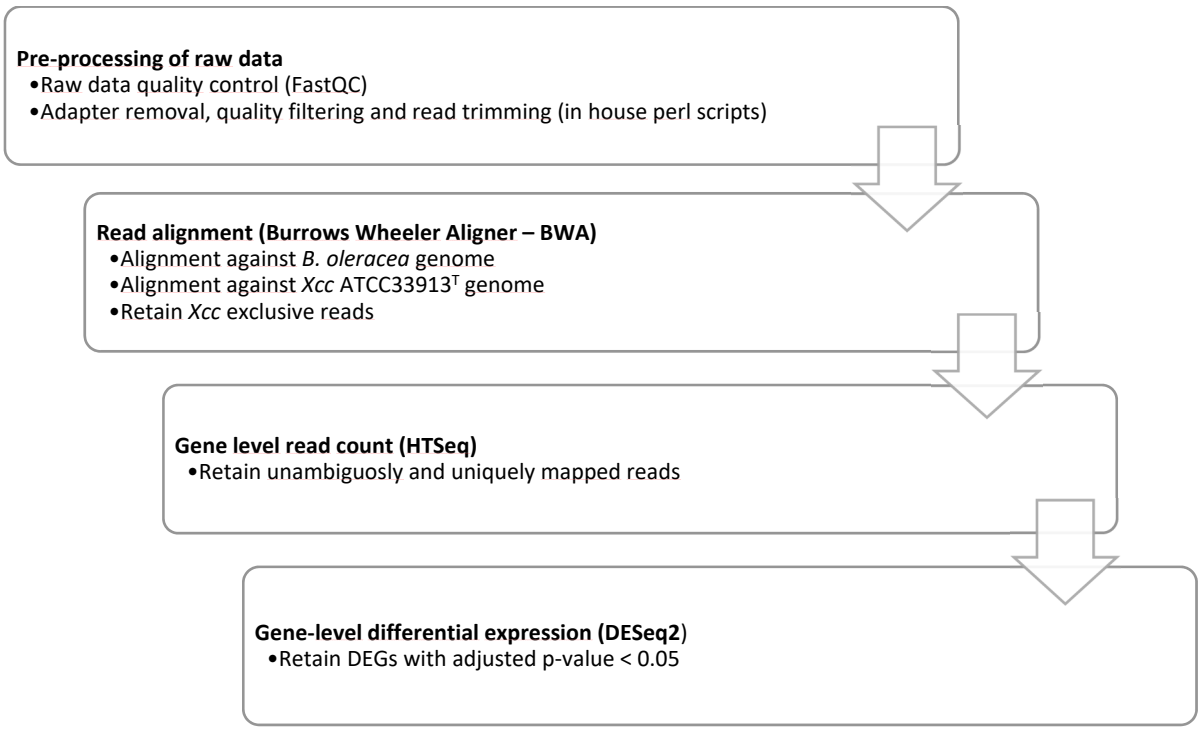


Figure 17 – Workflow for RNA-Seq raw data analysis.

After an initial quality control of the raw reads generated from the high throughput sequencing, performed using FastQC (Babraham Bioinformatics), as exemplified in Figure 18, sequences were trimmed to a final size of 69 bp to remove sequence portions with a biased nucleotide balance (12bp and 1bp trimmed in the 5' and 3' end, respectively) and the filters of sequence quality were applied to improve alignment accuracy (removal of sequences corresponding to homopolymers and/or with an average quality score lower than 20), using an *in house* developed PERL script.

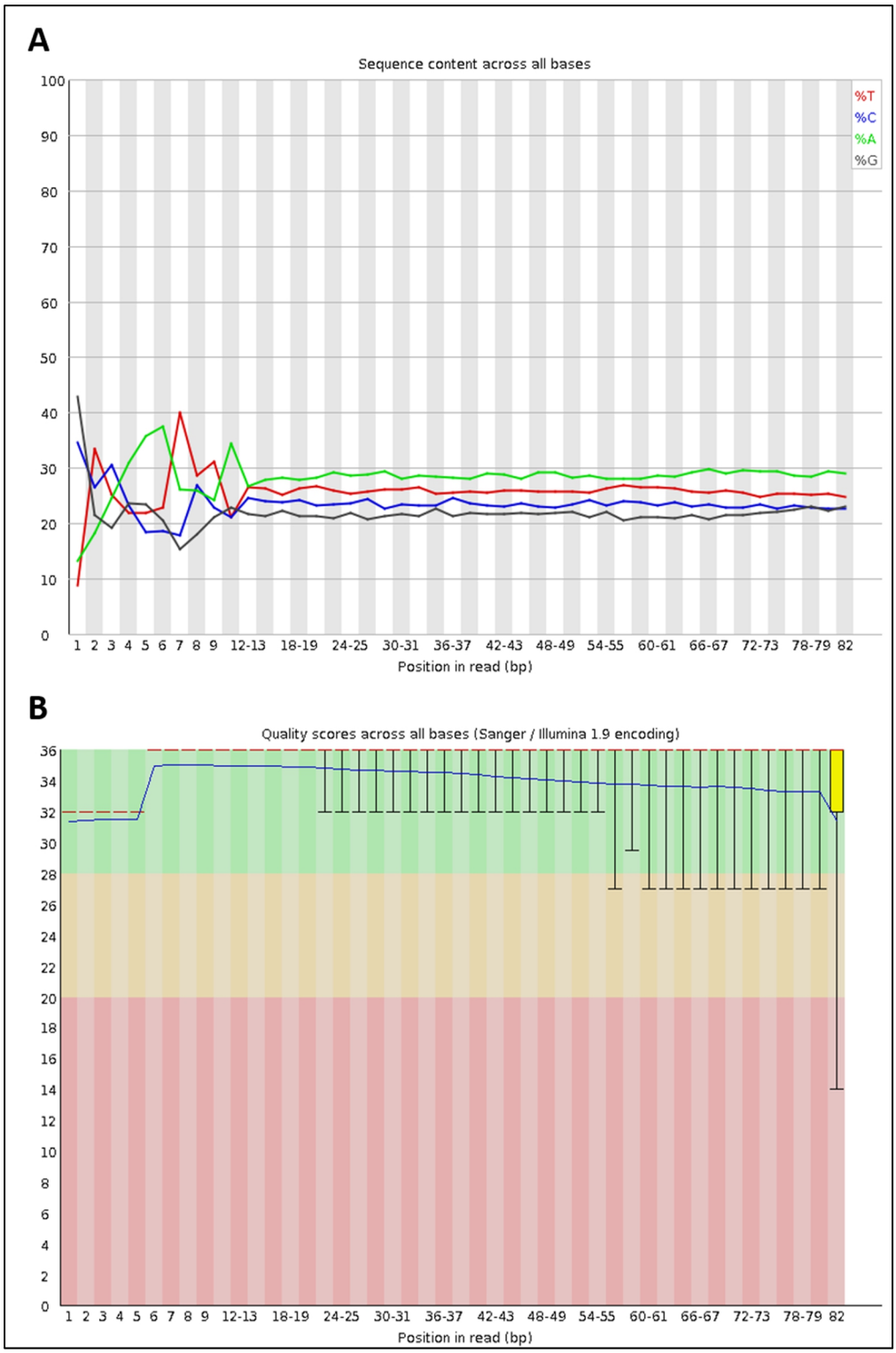


Figure 18 – FastQC quality control for sample WiL2 (*Brassica oleracea* var. *sabauda* cv. 'Wirosa F1'x *Xanthomonas campestris* pv. *campestris* strain L-vir [replicate 2]). (A) Per base sequence content, evidencing biased nucleotide balance in the first 12-13 bases; (B) Per base sequence quality, evidencing a quality score < 20 for base in position 82.

The remaining reads were mapped to the chromosomal and mitochondrial genome of *Brassica oleracea* (NCBI accessions NC_027748 -NC_027756; NC_016118) by using Burrows-Wheeler Aligner (BWA), a software package for mapping low-divergent sequences against a large reference genome (Li & Durbin, 2010). Up to three bases mismatches per read were allowed for mapping (`bwa aln -n 3 -t 4`). The reads that were not mapped in the *B. oleracea* genome were then mapped

against in the complete genome of *Xcc* type strain ATCC33913^T (NCBI accession NC_003902.1), using BWA and allowing for up to five bases mismatches per read (`bwa aln -n 5 -t 4`). To filter out any chloroplastidial sequences, the reads that mapped in the *Xcc* genome were then mapped in the chloroplastidial genome of B. oleracea (NCBI accession KR233156), using BWA and allowing up to five mismatches (`bwa aln -n 5-t 4`).

Gene counts were estimated using HTseq-count, a Phyton-based script designed specifically for differential expression analysis (Anders *et al.*, 2015). Using this script, only reads mapping unambiguously to a single gene were counted, whilst reads aligned to multiple positions or overlapping with more than one gene were discarded.

The obtained read counts per gene were then normalized using a scaling factor, in order to consider the varying sequencing depths of the different samples using DESeq2 R package 1.10.1 from Bioconductor (Love et al., 2014). At this stage, a data quality assessment and quality control was carried out, by exploring count and distance matrices using heat-maps, as well as by performing principal component analysis of the samples. This approach allowed to visualize the data and to remove sequences whose experimental treatment suffered from an abnormality, rendering them detrimental to the differential expression analysis. Finally, statistical testing of differential gene expression (DEG) was performed between all pairs of the studied conditions using the DESeq2 R package 1.10.1 from Bioconductor (Love *et al.*, 2014). Using DESeq2, the read counts per gene are modeled using a negative binomial distribution, modeling the observed relationship between the mean and variance when estimating dispersion, allowing a balanced selection of differentially expressed genes throughout the dynamic range of the data. The Benjamin–Hochberg procedure was used to control the false discovery rate (FDR).

2.3 Results and discussion

2.2.1 Total RNA purification, quantification and quality assessment

Total RNA concentration varied between 123-532 ng/µL, while RNA integrity number (RIN) varied between 7.0-8.3, as described in Table 12.

Table 12 – Total RNA sample concentration and RNA Integrity Number (RIN) of the eight samples used for RNA-Seq.

Pathosystem				
***Xcc* strain**	**Host**	**Sample ID**	**RNA (ng/µl)**	**RIN**
CPBF 278 'H-vir' (H)	*Brassica oleracea* convar. *acephala* var. *sabellica* cv. 'Bonanza F1' (Bo)	BoH1	480	7.6
		BoH2	123	7.9
CPBF 213 'L-vir' (L)	*Brassica oleracea* convar. *acephala* var. *sabellica* cv. 'Bonanza F1' (Bo)	BoL1	532	7.0
		BoL2	497	7.8
CPBF 278 'H-vir' (H)	*Brassica oleracea* var. *sabauda* cv. 'Wirosa F1' (Wi)	WiH1	160	8.1
		WiH2	174	8.3
CPBF 213 'L-vir' (L)	*Brassica oleracea* var. *sabauda* cv. 'Wirosa F1' (Wi)	WiL1	521	8.1
		WiL2	450	7.6

Bioanalyzer profiles were as expected for uncorrupted samples and did not reveal the presence of genomic DNA contamination, as evident in Figure 19A,B. RNA from all samples was successfully used to amplify a fragment of the 18S rRNA gene after RT-PCR, thus guaranteeing its applicability for downstream expression analysis, as depicted in Figure 19C.

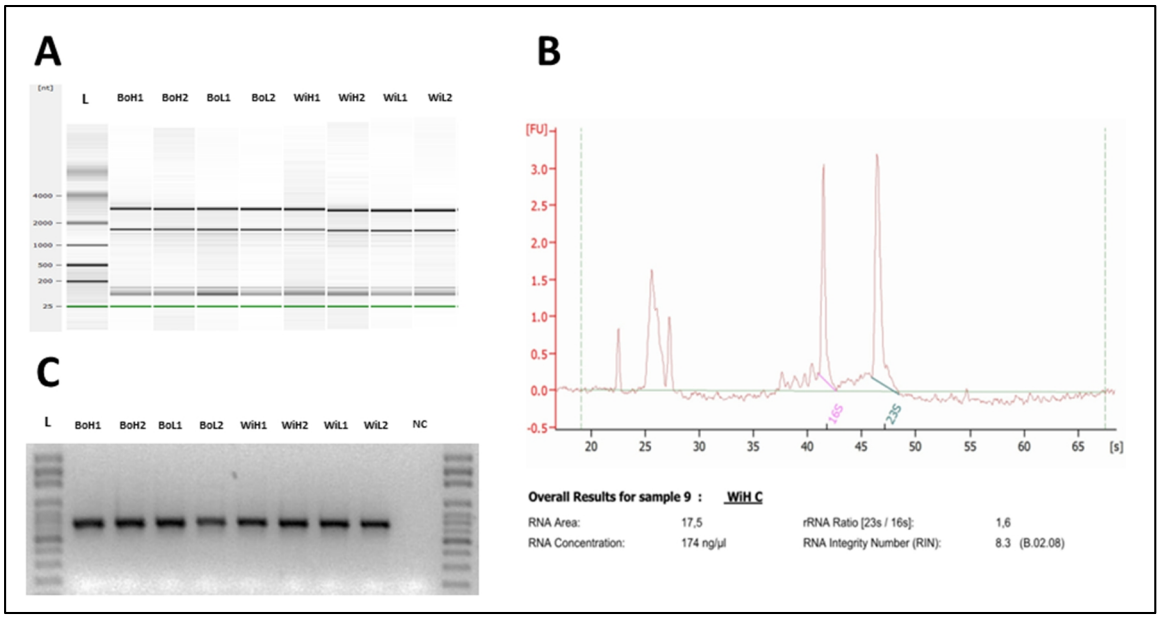


Figure 19 - Total RNA sample integrity. (A) Gel-like picture. (B) Electropherogram for sample WiH2 (*Brassica oleracea* var. *sabauda* cv. 'Wirosa F1'x *Xanthomonas campestris* pv. *campestris* strain H-vir [replicate 2]). (C) Functionality of RNA samples, assessed through PCR amplification of a 663bp fragment of the *Brassica oleracea* 18SrRNA gene after RT-PCR (NC – negative control).

2.2.2 RNA-Seq *in vivo* transcriptome profiling of *Xcc* – *Brassica oleracea* pathosystems

The RNA sequencing of the two replicates for each of the 4 established pathosystems yielded a total of 557M reads, resulting in 530M filtered reads with 69bp in length Table 13. Of the total filtered reads, an average of 65% mapped to the *B. oleracea* genome, while only 0.9% mapped to the *Xcc* genome. The remaining 34% of reads were not mapped against any of these organisms.

Table 13 - Summary of Illumina sequencing data.

	Bonanza x H-vir		Bonanza x L-vir		Wirosa x H-vir		Wirosa x L-vir	
	BoH1	BoH2	BoL1	BoL2	WiH1	WiH2	WiL1	WiL2
Raw reads	64,627,709	70,164,498	57,684,189	60,101,232	83,776,418	74,897,476	83,768,596	62,095,380
Filtered reads	61,709,268	67,238,668	55,298,072	56,704,063	78,751,750	71,723,704	79,927,590	58,849,364
B.oleracea	40,889,642 (66%)	44,813,469 (67%)	35,467,470 (64%)	38,942,300 (69%)	49,557,391 (63%)	45,399,893 (63%)	49,663,919 (62%)	38,095,887 (65%)
Xcc	1,119,409 (1,8%)	826,650 (1,2%)	215,722 (0,4%)	383,478 (0,7%)	462,984 (0,6%)	1,077,414 (1,5%)	440,236 (0,6%)	274,041 (0,5%)
Not aligned	19,700,217 (32%)	21,598,549 (32%)	19,614,880 (35%)	17,378,285 (31%)	28,731,375 (36%)	25,246,397 (35%)	29,823,435 (37%)	20,479,436 (35%)

Unmapped reads are common in RNA-Seq experiments, and are most likely the result of divergences between the reads and the reference genomes used for mapping, a phenomenon known as reference genome bias (Sousa & Hey, 2013). In this study, due to the unavailability of genomes for the varieties used, reads were mapped against a different B. oleracea reference genome, which could result in mapping divergences. Significant divergence between genomes has been reported for closely related genomes of other plants. In a comparative analysis of two lines of Oryza sativa subsp. indica, several structural differences were found, including inversions, translocations, presence/absence variations, and segmental duplications (Zhang et al., 2016). Similarly, considering the genomic diversity described for Xcc (da Silva et al., 2002; Qian et al., 2005), it is also likely that some Xcc reads failed to map against the reference genome used.

In addition to reference genome bias, drawbacks resulting of contamination or errors during library preparation can increase the number of unmapped reads (Gouin *et al.*, 2015). Moreover, and although the assays were carried out under controlled conditions, the presence of reads from other organisms cannot be 100% ruled out, and can thus contribute to a higher proportion of unmapped reads.

Although in most studies the unmapped reads are discarded, in favor of mapped reads, a *de novo* assembly of such read set could improve mapping results. In fact, in a similar strategy was successfully used in a dual RNA-Seq transcriptional analysis of wheat colonized by the plant growth promoting bacterium *Azospirillum brasilense* (Camilios-Neto *et al.*, 2014).

Xcc mapping reads ranged from 215722 (sample BoL1) to 1119409 (sample BoH1), corresponding to a genome coverage ranging from 3 to 15 times (Table 13). Low coverage of *Xcc* genome is the direct result of the low abundance of bacterial transcripts in the original RNA samples.

Mapping at gene level ranged from 5984 reads to 43425 reads for samples BoL1 and BoH1, respectively, corresponding to a total average of 3% of *Xcc* reads. Low mapping at gene level can be due to several factors, such as presence of antisense reads, ambiguous, low quality reads and reads mapping to regions not described as genes in the reference genome annotation. The latter include reads mapping to regions identified in the annotation as rRNA, tRNA, transcript, region, protein-binding sites or even intergenic regions, which represent 15% of the annotation genome of *Xcc* ATCC33913^{T}. Although the genome used for read alignment corresponds to the reference genome for the species, mapping to gene level could be improved by using a strain with a more thoroughly annotated genome, such as *Xcc* strain 8004.

Genome browsing of the analyzed samples using the Integrative Genomics Viewer tool (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) also revealed the presence of rRNA mapped reads, which suggests that the rRNA depletion step during library preparation was not efficient, most likely due to inappropriate probe specificity. rRNA depletion step was conducted at EMBL GeneCore facilities using RiboZero kit, according to the specifications recommended by Illumina, with a mixture of plant and bacterial rRNA probes. Although these probes have been tested from some plant-only and bacteria-only samples, for *Brassica oleracea* and *Xanthomonas campestris*, no *in silico* compatibility is reported by the kit manufacturer. Similarly, there is no available information on the behavior of mixed rRNA probes for the depletion of mixed samples.

The effects of rRNA depletion on transcriptome profiling have been reported in a comprehensive study on the use of RNA-Seq to study bacterial transcriptomes, using *E. coli* as a model (Haas et al., 2012). In that study, 82% of the fragments in an undepleted sample aligned to rRNAs, although there was still at least one fragment mapped to over 99% of annotated ORFs. Therefore, the lack of rRNA depletion reduced the number of mRNA reads in the library, but only led to a modest decrease in the proportion of annotated ORFs detected. In fact, the authors found that the number of reads obtained for undepleted samples was sufficient for differential expression studies.

This has also been reported in previous studies on the transcriptome of *Xcc*: in a work comparing the transcriptomes of *Xcc* strain 8004 grown in two different culture media, only 50% of the rRNA depleted libraries mapped to genes. The remaining reads were mapped to non-coding regions, as well as to rRNA genes, tRNA genes, tmRNA gene, intergenic regions, insertion elements, the gene for RNA subunit of RNase P and several multiple copy genes (Liu et al., 2013). Still, the authors were able to find expression for 4220 of the 4297 genes annotated in *Xcc* strain 8004 genome, in a total of 92% of the genes (Liu et al., 2013; Qian et al., 2005).

The negative outcome of rRNA presence in the analyzed libraries is associated with the absence of low-abundance transcripts. Despite these constraints, the fact that the libraries contain rRNA ensures that only the most expressed genes were sampled in the subsequent steps of gene-level counts and expression analysis.

The results presented in this thesis are in line with what was described by Liu et al (2013): although reads deriving from non-mRNA fragments are present, expression was reported for 3866 of the 4240 genes annotated in *Xcc* ATCC33913^{T} genome, in a total of 91% of the genes.

In the four pathosystems analyzed, the number of expressed genes varied from 2570 for WiL sample to 3576 to BoH sample, as highlighted in Figure 20A. Samples obtained from pathosystems containing *Xcc* L-vir strain recorded a lower number of expressed genes, when compared to pathosystems containing *Xcc* H-vir strain, which could be the result of a lower bacterial population in L-vir pathosystems. The number of expressed genes in samples from pathosystems containing the same strain was equivalent: 2570 vs. 2643 for WiL and BoL, respectively; and 3397 vs. 3576 for WiH and BoH, respectively. Of the 4240 genes annotated in *Xcc* ATCC33913T genome, 1899 were expressed in all four samples. Considering the pooled samples obtained from cvs. Bonanza and Wirosa, a total of 3456 Xcc genes were expressed in both conditions, while 261 were exclusively expressed in cv. Bonanza and 149 were exclusively expressed in cv. Wirosa (Figure 20B). Concerning the pooled samples containing H-vir strain against those containing L-vir strain, a total of 3050 genes were expressed by both strains, while 695 were exclusively expressed in H-vir and 121 in L-vir (Figure 0C).

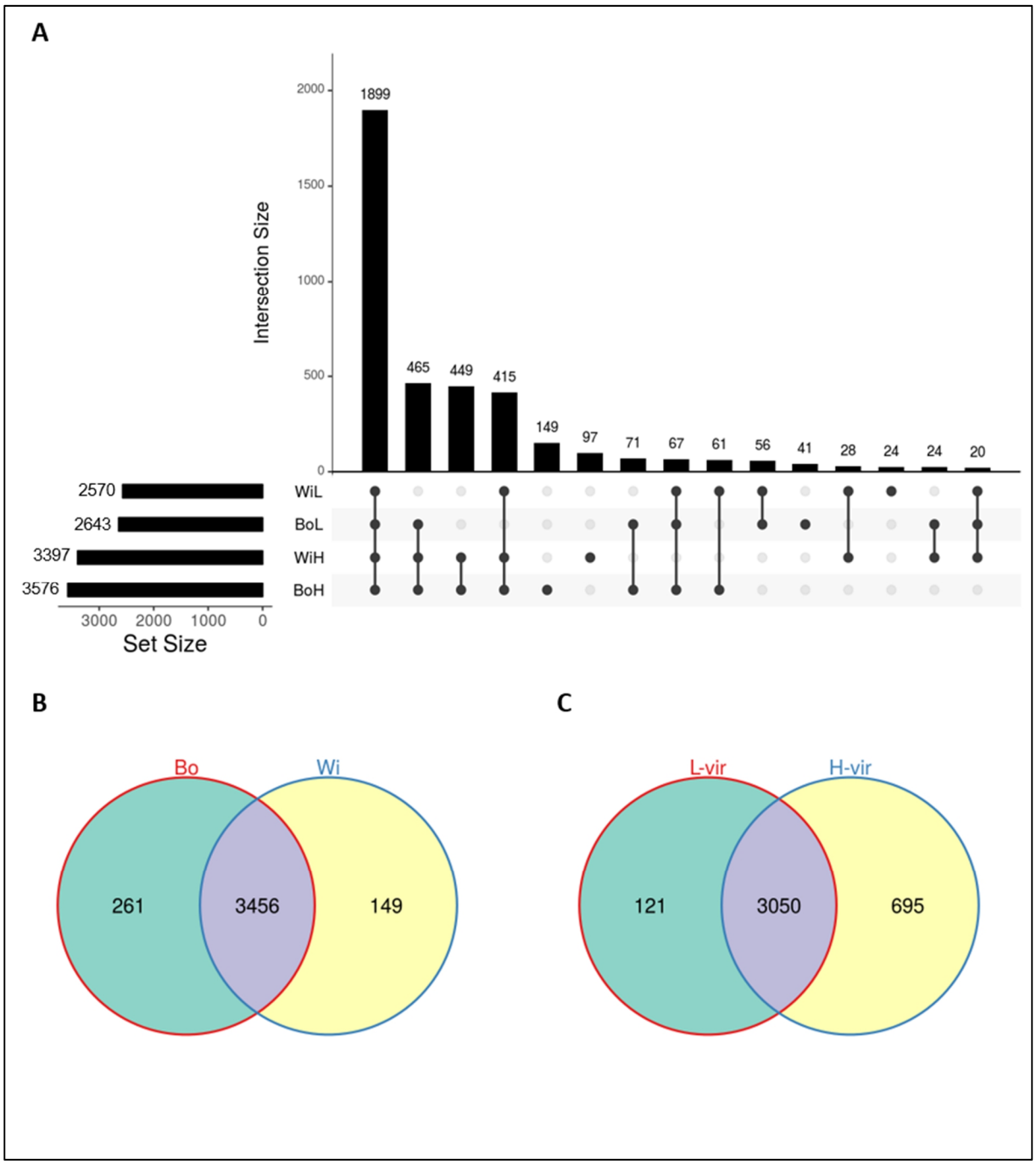


Figure 20 – *Xanthomonas campestris* pv. *campestris* (*Xcc*) gene expression. (A) Intersection of *Xcc* expressed genes (EGs) in analyzed pathosystems *Brassica oleracea* cv. *Wirosa* x L-vir strain (WiL), *B. oleracea* cv. *Bonanza* x L-vir strain (BoL), *B. oleracea* cv. *Wirosa* x H-vir strain (WiH) and *B. oleracea* cv. *Bonanza* x H-vir strain (BoH) (calculated using the total number of EGs from 2 replicate samples). Set size refers to the total number of EGs in each pathosystem. B. *Xcc* EGs in samples obtained from Bonanza and Wirosa. C. *Xcc* EGs in samples containing L-vir and H-vir. UpSet plot and Venn diagrams were created using the UpSet and Venn modules of Intervene Shiny App (Khan & Mathelier, 2017).

2.2.3 *In vivo* differential gene expression of Low-virulence and High-virulence *Xcc* strains

Exploratory data quality assessment and visualization, using heat-maps of sample distances suggested that no differences among samples could be attributed to different hosts (Figure 21A). Likewise, principal component analysis, retaining 72% of total sample variance, did not reveal any grouping based on the host (Figure 21B). In fact, these approaches revealed a tendency of sample clustering based solely on the *Xcc* strain used in each pathosystem. Additionally, when performing differentially expressed gene (DEG) analysis for each separate strain, no gene was found differentially expressed for any of the hosts tested (Figure 21C, D). For this reason, the subsequent statistical analyses were carried out using the four samples from each strain, regardless of the host of origin, and after filtering out genes with low expression. This analysis confirmed the previous findings, since there was a clustering of the samples by strain (Figure 22A, B). Principal component analysis, retaining 73% of total sample variance, also confirmed this grouping (Figure 22C). Using DESeq2 package to analyze L-vir vs. H-vir, a total of 154 DEGs were identified, with log_2FC values ranging from -7,60 to 7,17. Of the 154 DEGs, 64 were up-regulated and 90 were down-regulated in L-vir strain (Figure 22D).

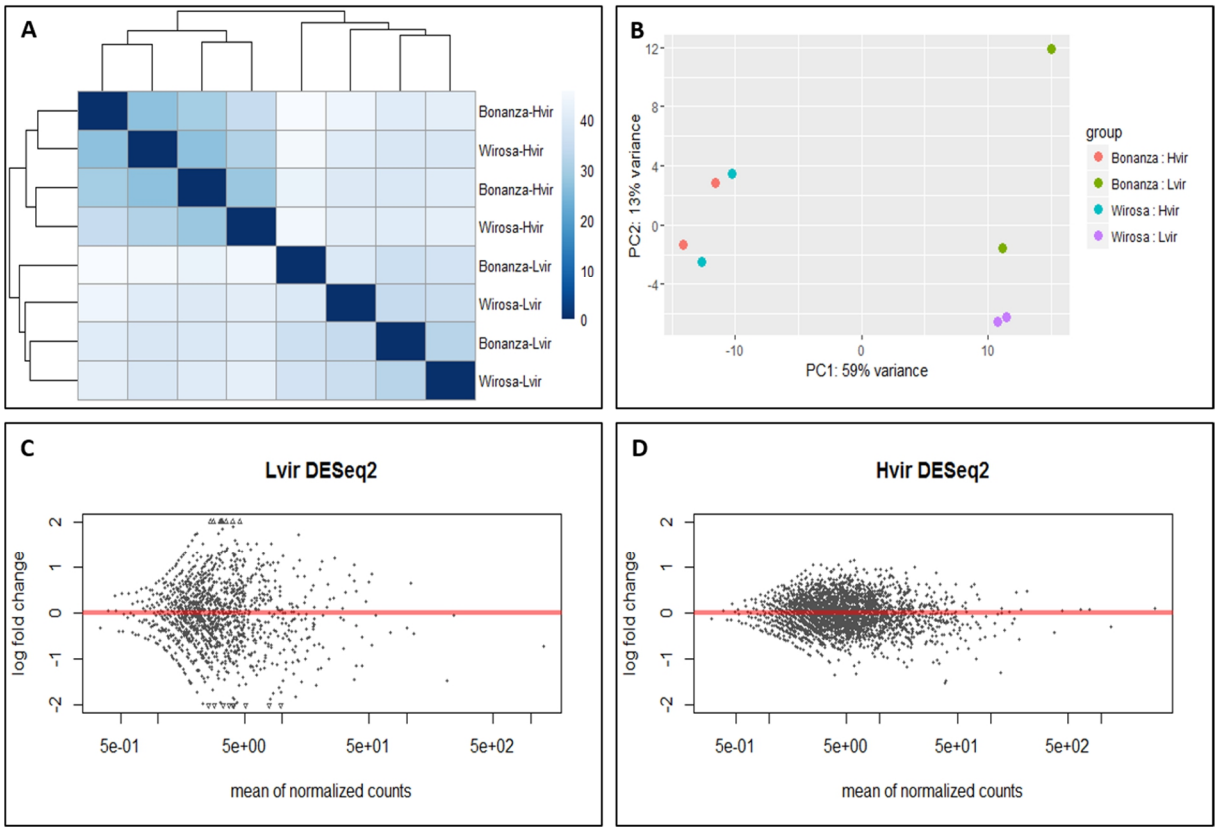


Figure 21 - Exploratory analysis of RNA-Seq reads. (A) heat-map of sample to sample distances. (B) Principal Component Analysis (PCA) of samples. *Volcano* plots of differential gene expression between hosts (Wirosa vs. Bonanza), for L-vir (C) and H-vir (D).

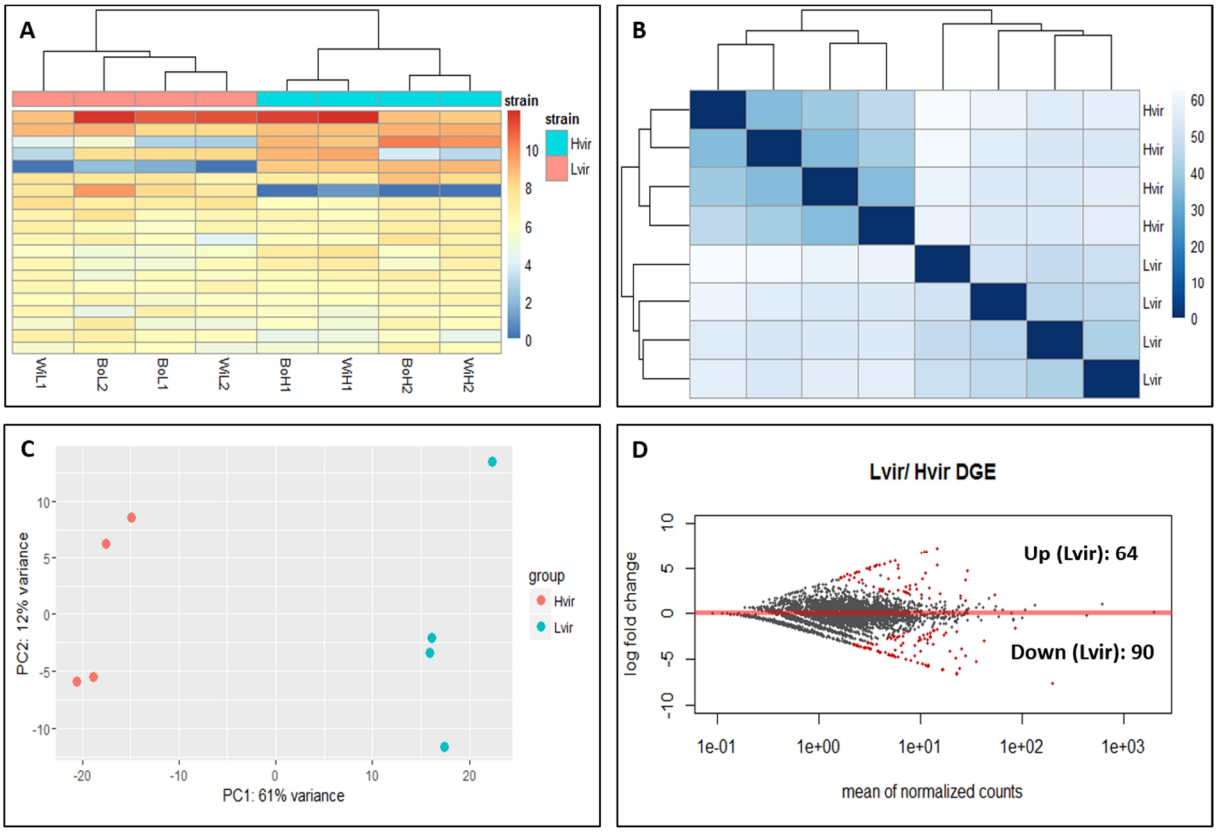


Figure 22 - Comparison of the transcriptomes of *Xanthomonas campestris* pv. *campestris* – *Brassica oleracea* pathosystems. (A) Clustering heat-map of the top 20 most expressed genes per sample (log_2 normalized counts). (B) Clustering heat-map of sample-to-sample distances. (C) Principal Component Analysis of High-virulence and Low-virulence samples. (D) *Volcano* plot of differential gene expression (DGE) between L-vir and H-vir strains (Lvir/Hvir DGE), red dots represent significant DEGs (adjusted p-value<0.05).

The 154 DEGs identified were assigned to 20 COG (Cluster of Orthologous Groups) categories according to the information available in the *Xanthomonas* Genome Browser (www.xgb.org) and NCBI Gene database (Figure 23). These categories included energy production and conversion (C); cell cycle control, mitosis and meiosis (D); amino acid transport and metabolism (E); nucleotide transport and metabolism (F); carbohydrate transport and metabolism (G); coenzyme transport and metabolism (H); lipid transport and metabolism (I); translation, ribosomal structure and biogenesis (J); transcription (K); replication, recombination and repair (L); cell wall/ membrane biogenesis (M); cell motility (N); posttranslational modification, protein turnover and chaperones (O); inorganic ion transport and metabolism (P); secondary metabolites biosynthesis, transport and catabolism (Q); general function prediction only (R); function unknown (S); signal transduction mechanisms (T); intracellular trafficking and secretion (U) and defense mechanisms (V). Additionally, there were poorly characterized genes with unclassified COG category and genes that were not analyzed (such as tRNA genes). Assessment of the relative abundance of each COG category in each set of up and down-regulated genes, showed that 61% of the genes have known function, while the remaining 39% of the genes are poorly characterized, falling under COG categories R and S as well as 'unclassified' and 'not analyzed' (Figure 23).

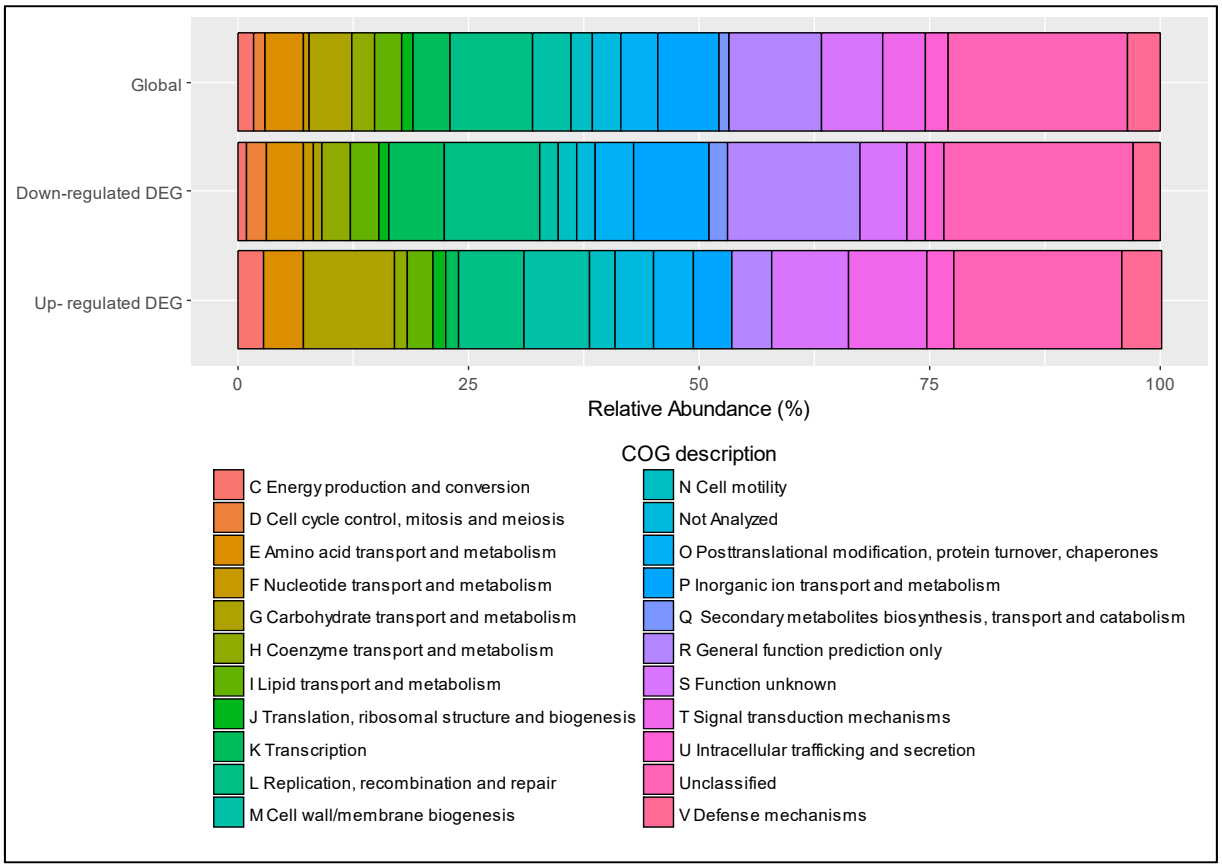


Figure 23 - Relative abundance of COG (Cluster of Orthologs Groups) categories among the global differentially expressed genes (DEG) set, the down-regulated DEG set and the up-regulated DEG set.

COG description for *Xcc* did not allow to assess if a given biological pathway is induced or repressed in *Xcc* during pathogenesis. To address this issue, each DEG is described below with further detail, emphasizing their biological role within the cell. To allow a better integration of the results obtained in this study with what has been previously described for this pathogen, the gene categories defined by He *et al.* (2007) have been maintained, similarly to what was performed by Liu *et. al* (2013). Therefore, maintaining the COG description as reference, a more accurate description of biological role is achieved. For COG poorly characterized genes, the information obtained from other sources (KEGG pathways, protein motifs and previous reports) was used, when available, to allocate them into the most appropriate category.

This approach allowed improving the characterization of the identified DEGs, consequently reducing the proportion of poorly characterized genes from 39% (based solely on COG description) to 27%, as depicted in Figure 24A. Of the remaining 73% of the global differentially expressed gene set, the most represented category was ‘Pathogenicity and adaptation’.

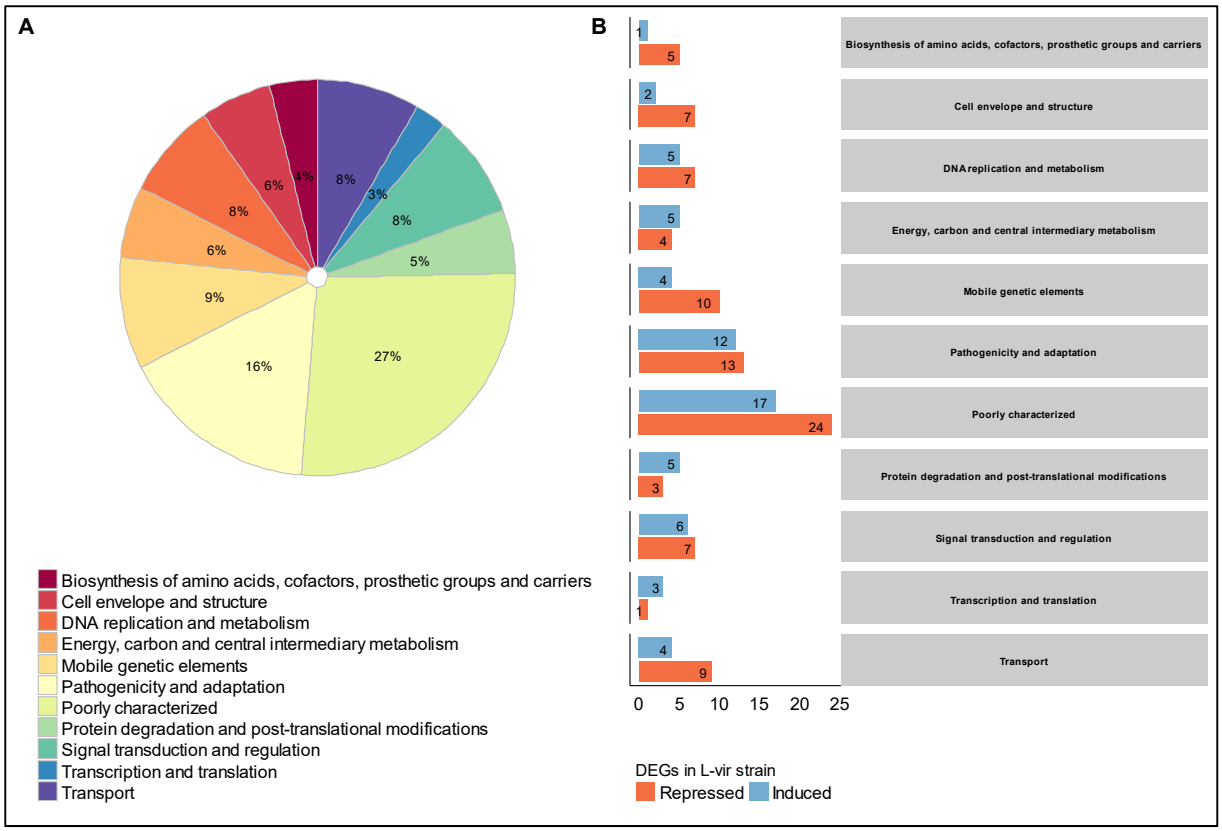


Figure 24 – (A) Relative abundance of differentially expressed genes according to their biological role in the cell. (B) Number of repressed and induced differentially expressed genes (DEGs) found in *Xanthomonas campestris* pv. *campestris* Low-virulence strain, according to their biological role.

Despite the difference in the number of genes induced or repressed in the least virulent strain, no category appears to be favored for induction or repression (Figure 24B). In fact, the most abundant well characterized category - 'Pathogenicity and adaptation' - comprises nearly the same number of repressed and induced genes.

The identified DEGs are listed in Table 14, along with their annotation based on the genome of type strain ATCC33913^T, COG category, biological role and log_2FC value.

Table 14 – Differentially expressed gene ID, name (when existent), annotation, COG (Cluster of Ortholog Groups) description and Biological role. log_2FC calculated using L-vir/H-vir gene counts.

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
Down-regulated genes in L-vir strain								
XCC3695		oxidoreductase	IQR	Energy, carbon and central intermediary metabolism	1,35	395,03	-7,60	4,64E-25
XCC0536		hypothetical protein	Unclassified	Poorly characterized	0,00	45,16	-6,52	7,54E-09
XCC1461	*repA*	phage-related protein	J	Mobile genetic elements	0,00	45,73	-6,51	1,60E-08
XCC3340		glycosyl transferase family protein	M	Cell envelope and structure	0,00	33,67	-6,14	7,16E-08
XCC2565	*xopAC*	leucin rich protein	S	Pathogenicity and adaptation	0,00	32,51	-6,06	2,04E-07
XCC0670	*xpsD*	general secretion pathway protein D	NU	Pathogenicity and adaptation	0,29	48,30	-5,95	1,69E-07
XCC1463		hypothetical protein	K	Transcription and translation	0,00	23,95	-5,66	1,80E-06
XCC3989		hypothetical protein	Unclassified	Poorly characterized	0,00	23,60	-5,65	2,15E-06
XCC3935		tRNA-Thr_NC_003902.1	Not analyzed	Poorly characterized	0,00	24,07	-5,58	1,26E-05
XCC0538		hypothetical protein	Unclassified	Poorly characterized	0,00	22,39	-5,55	4,48E-06
XCC3584		hypothetical protein	Unclassified	Poorly characterized	0,00	20,82	-5,48	4,52E-06
XCC3932		transcriptional regulator (AraC family?)	K	Signal transduction and regulation	0,00	18,89	-5,34	1,26E-05
XCC0055		hypothetical protein	S	Poorly characterized	1,69	69,58	-5,17	4,01E-10
XCC0030		hypothetical protein	Unclassified	Poorly characterized	0,00	16,13	-5,07	9,99E-05
XCC0526		hypothetical protein	R	Poorly characterized	0,00	14,49	-4,96	1,10E-04
XCC1846	*recJ*	single stranded DNA exonuclease	L	DNA replication and metabolism	1,35	43,10	-4,78	5,14E-07
XCC0491		hypothetical protein	EP	Transport	0,00	11,94	-4,73	2,34E-04
XCC4114		hypothetical protein	Unclassified	Poorly characterized	0,00	11,31	-4,67	2,69E-04
XCC1643	*gtrB*	glycosyl transferase-like protein	R	Cell envelope and structure	0,00	10,89	-4,60	5,23E-04
XCC2816		IS1480 transposase	L	Mobile genetic elements	0,29	17,10	-4,59	1,93E-04
XCC1633		ISxac3 transposase	L	Mobile genetic elements	0,00	11,10	-4,58	7,25E-04

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC0507		ice nucleation protein	Unclassified	Pathogenicity and adaptation	0,00	10,41	-4,55	5,55E-04
XCC4115	*xynA*	endo-1,4-beta-xylanase A (ring canal kelch-like protein)	Unclassified	Pathogenicity and adaptation	1,93	53,14	-4,53	3,02E-08
XCC2692	*creD*	hypothetical protein (inner membrane protein)	V	Cell envelope and structure	0,00	9,84	-4,50	5,89E-04
XCC0258	*xopR*	hypothetical protein	D	Pathogenicity and adaptation	0,00	9,74	-4,47	5,89E-04
XCC0537		hypothetical protein	Unclassified	Poorly characterized	0,00	10,57	-4,45	1,76E-03
XCC2904		type I restriction enzyme specificity chain-like protein	V	DNA replication and metabolism	0,00	9,09	-4,27	3,47E-03
XCC1303		hypothetical protein	Unclassified	Poorly characterized	0,29	12,81	-4,22	1,03E-03
XCC2867	*btuB*	TonB-dependent receptor	P	Transport	0,00	8,12	-4,22	1,69E-03
XCC4141		ISxac3 transposase	L	Mobile genetic elements	0,00	7,99	-4,20	1,93E-03
XCC0304		TonB-dependent receptor	P	Transport	0,00	7,73	-4,14	2,48E-03
XCC1669		cytochrome like B561	C	Energy, carbon and central intermediary metabolism	0,78	19,43	-4,14	7,10E-04
XCC2076		hypothetical protein (phage-related protein)	Unclassified	Mobile genetic elements	0,00	7,80	-4,11	4,66E-03
XCC3934		hypothetical protein	R	Poorly characterized	0,00	7,57	-4,11	3,58E-03
XCC3734		tRNA-Met_NC_003902.1	Not analyzed	Poorly characterized	0,00	7,77	-4,09	5,96E-03
XCC1930		hypothetical protein	QR	Poorly characterized	1,86	37,14	-4,05	1,09E-06
XCC2366		TetR/AcrR family transcriptional regulator	K	Signal transduction and regulation	0,46	10,76	-3,98	2,37E-03
XCC0524		helicase	K	DNA replication and metabolism	0,00	6,50	-3,88	1,04E-02
XCC1133		hypothetical protein	Unclassified	Poorly characterized	0,00	6,30	-3,86	1,09E-02
XCC0355	*pobR*	PobR regulator (AraC family)	Unclassified	Signal transduction and regulation	2,71	47,58	-3,84	1,27E-07
XCC1134		hypothetical protein	S	Poorly characterized	0,00	5,79	-3,78	9,90E-03
XCC4191	*cII*	phage-related regulatory protein cII	S	Mobile genetic elements	0,00	6,04	-3,76	1,64E-02
XCC0136		IS1480 transposase	D	Mobile genetic elements	0,00	5,82	-3,75	1,36E-02
XCC2547		hypothetical protein	L	Poorly characterized	0,00	5,70	-3,74	1,09E-02

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC2399		hypothetical protein	R	Transport	0,00	5,55	-3,72	1,16E-02
XCC4194		hypothetical protein	Unclassified	Poorly characterize	0,00	5,25	-3,61	2,15E-02
XCC3363		hypothetical protein	V	DNA replication and metabolism	0,00	4,95	-3,56	2,15E-02
XCC0790	*apaG*	ApaG protein	P	Transport	0,00	4,84	-3,52	2,32E-02
XCC3626		RNA-directed DNA polymerase	L	DNA replication and metabolism	0,00	4,88	-3,52	2,23E-02
XCC0850	*ilvE*	branched-chain amino acid aminotransferase	L	Biosynthesis of amino acids, cofactors, prosthetic groups and carriers	0,95	11,84	-3,52	3,95E-03
XCC2538		oxidoreductase	EH	Energy, carbon and central intermediary metabolism	1,23	16,33	-3,49	1,95E-03
XCC1773		hypothetical protein	R	Poorly characterized	0,00	4,71	-3,49	2,98E-02
XCC1976		recombination factor protein RarA	Unclassified	DNA replication and metabolism	0,00	4,72	-3,45	2,90E-02
XCC2172	*cpo*	non-heme chloroperoxidase	L	Pathogenicity and adaptation	0,37	7,05	-3,40	2,15E-02
XCC4075		hypothetical protein (Putative manganese efflux pump MntP)	R	Transport	0,00	4,42	-3,40	3,55E-02
XCC0257		nitrile hydratase activator	S	Signal transduction and regulation	2,35	24,60	-3,38	3,73E-04
XCC2395	*btuB*	TonB-dependent receptor	R	Transport	0,00	4,38	-3,37	3,84E-02
XCC4002		hypothetical protein	P	Poorly characterized	0,29	6,56	-3,32	2,66E-02
XCC0930		hypothetical protein (Cell wall hydrolase, SleB?)	R	Cell envelope and structure	0,00	4,12	-3,29	4,91E-02
XCC2053	*cheY*	response regulator	M	Signal transduction and regulation	2,05	21,57	-3,26	5,32E-04
XCC3980	*apbA*	2-dehydropantoate 2-reductase	T	Biosynthesis of amino acids, cofactors, prosthetic groups and carriers	0,37	6,06	-3,19	3,78E-02
XCC3041		hypothetical protein	H	Poorly characterized	0,37	6,13	-3,18	4,51E-02
XCC0191	*sodC2*	superoxide dismutase	Unclassified	Pathogenicity and adaptation	0,29	5,84	-3,17	4,15E-02
XCC1941	*fliC*	flagellin	P	Pathogenicity and adaptation	1,03	12,32	-3,17	5,59E-03
XCC0400		IS1404 transposase	N	Mobile genetic elements	2,05	17,30	-2,96	1,93E-03
XCC0530	*xopX*	hypothetical protein	L	Pathogenicity and adaptation	9,37	75,58	-2,90	7,26E-07

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC4050		ankyrin-like protein	Unclassified	Cell envelope and structure	2,29	17,88	-2,81	1,76E-03
XCC0142		hypothetical protein (hydrolase activity)	R	Biosynthesis of amino acids, cofactors, prosthetic groups and carriers	1,32	10,00	-2,80	2,39E-02
XCC0342		xanthomonadin biosynthesis related protein	E	Pathogenicity and adaptation	2,80	20,56	-2,79	8,49E-04
XCC3733		ISxac3 transposase	Unclassified	Mobile genetic elements	1,11	9,16	-2,69	4,60E-02
XCC0221		acetyltransferase	L	Protein degradation and post-translational modifications	2,80	19,10	-2,67	5,86E-03
XCC3187	*idgB*	indigoidine synthesis like protein	R	Energy, carbon and central intermediary metabolism	1,03	8,43	-2,67	4,28E-02
XCC2810	*mgtE*	Mg^{2+} transporter	R	Transport	2,66	16,86	-2,64	5,96E-03
XCC2766	*ynhE*	cysteine desulfurase (ABC transporter permease)	P	Transport	3,04	21,81	-2,61	1,59E-03
XCC2277	*suhB*	extragenic supressor protein	O	Signal transduction and regulation	1,40	10,40	-2,61	2,34E-02
XCC2386	*upp*	uracil phosphoribosyltransferase (UMP phosphotransferase)	G	DNA replication and metabolism	1,93	10,37	-2,48	3,84E-02
XCC2359		hypothetical protein	F	Poorly characterized	1,49	9,49	-2,45	3,92E-02
XCC3961		dipeptidyl peptidase IV	S	Protein degradation and post-translational modifications	1,80	11,68	-2,45	2,15E-02
XCC2077	*trbP*	conjugal transfer protein TrbP	E	Mobile genetic elements	3,49	21,31	-2,41	1,68E-03
XCC0177	*cls*	cardiolipin synthase	Unclassified	Cell envelope and structure	1,64	9,27	-2,40	4,36E-02
XCC3324	*ilvG*	acetolactate synthase 2 catalytic large subunit	I	Biosynthesis of amino acids, cofactors, prosthetic groups and carriers	3,12	18,17	-2,35	1,19E-02
XCC2133		hypothetical protein (Putative aminomethyl transferase)	EH	Biosynthesis of amino acids, cofactors, prosthetic groups and carriers	4,98	29,26	-2,34	1,35E-03
XCC2715		hypothetical protein	R	Poorly characterized	3,89	19,74	-2,25	9,80E-03
XCC0006		hypothetical protein	O	Protein degradation and post-translational modifications	6,46	29,46	-2,13	5,96E-03

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC3936		hypothetical protein	Unclassified	Poorly characterized	4,19	20,66	-2,08	3,55E-02
XCC0407		hypothetical protein	Unclassified	Pathogenicity and adaptation	8,68	36,45	-1,98	5,89E-04
XCC0659		protease	T	Pathogenicity and adaptation	5,68	21,26	-1,84	1,24E-02
XCC3649	*lrp*	leucine responsive regulatory protein	O	Signal transduction and regulation	5,84	21,17	-1,80	3,64E-02
XCC2267		hypothetical protein	K	Cell envelope and structure	6,69	22,37	-1,67	2,06E-02
XCC0851	*prtA*	extracellular protease	I	Pathogenicity and adaptation	41,92	128,09	-1,58	3,08E-03
Up-regulated genes in L-vir strain								
XCC1629	*xopE2*	avirulence protein	Unclassified	Pathogenicity and adaptation	29,21	0,00	7,17	5,34E-10
XCC2095		DNA helicase-like protein	L	DNA replication and metabolism	22,67	0,00	6,85	2,98E-09
XCC3131		sensor kinase	T	Signal transduction and regulation	21,92	0,00	6,77	1,34E-08
XCC2055	*ndvB*	cyclic beta 1-2 glucan synthetase	G	Pathogenicity and adaptation	20,42	0,00	6,69	1,34E-08
XCC2096		hypothetical protein	Unclassified	Poorly characterized	11,37	0,00	5,87	4,38E-06
XCC0464		anticodon nuclease	S	Transcription and translation	10,05	0,00	5,68	1,44E-05
XCC2674		serine protease	O	Protein degradation and post-translational modifications	7,85	0,00	5,38	6,28E-05
XCC3144		hypothetical protein	K	Transcription and translation	8,00	0,00	5,37	5,79E-05
XCC2098		ISxcd1 transposase	L	Mobile genetic elements	11,91	0,09	5,34	1,26E-05
XCC2694	*creC*	sensory histidine kinase CreC	T	Signal transduction and regulation	28,95	0,51	5,34	2,88E-09
XCC2093		hypothetical protein	Unclassified	Poorly characterized	7,48	0,00	5,30	9,99E-05
XCC0605	*wxcE*	hypothetical protein WxcE	Unclassified	Pathogenicity and adaptation	7,10	0,00	5,23	9,57E-05
XCC2094		Tannase precursor	Unclassified	Poorly characterized	7,21	0,00	5,22	1,09E-04
XCC3146	*intS*	phage-related integrase	L	Mobile genetic elements	7,10	0,00	5,21	1,89E-04
XCC2732		hypothetical protein	V	DNA replication and metabolism	7,07	0,00	5,19	1,93E-04
XCC1027		hypothetical protein	Unclassified	Poorly characterized	6,29	0,00	5,07	3,15E-04

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC0462	*hsdR*	type I restriction-modification system endonuclease	V	DNA replication and metabolism	6,15	0,00	5,00	3,93E-04
XCC0606		GDP-mannose 4%2C6-dehydratase	M	Pathogenicity and adaptation	5,95	0,00	4,95	5,58E-04
XCC1894		hypothetical protein	Unclassified	Poorly characterized	12,06	0,00	4,88	1,46E-02
XCC2047	*tsr*	chemotaxis protein	NT	Signal transduction and regulation	56,32	1,77	4,73	1,35E-13
XCC2733		McrB-like protein	V	DNA replication and metabolism	5,00	0,00	4,66	2,49E-03
XCC0601	*wzt*	ABC transporter ATP binding protein	GM	Pathogenicity and adaptation	4,77	0,00	4,64	2,49E-03
XCC0604	*wxcD*	hypothetical protein	M	Pathogenicity and adaptation	8,78	0,00	4,60	2,44E-02
XCC2992	*T*	phage-related tail protein	S	Mobile genetic elements	4,53	0,00	4,52	2,70E-03
XCC2896	*xopD*	virulence protein	Unclassified	Pathogenicity and adaptation	21,47	0,87	4,40	1,69E-07
XCC0608	*rmd*	UDP-glucose 4-epimerase	GM	Pathogenicity and adaptation	4,09	0,00	4,34	9,34E-03
XCC3129	*virB6*	VirB6 protein	U	Pathogenicity and adaptation	3,82	0,00	4,32	5,86E-03
XCC3729		response regulator	T	Signal transduction and regulation	3,54	0,00	4,20	1,23E-02
XCC0609		IS1479 transposase	L	Mobile genetic elements	3,37	0,00	4,09	2,15E-02
XCC3130		hypothetical protein	Unclassified	Poorly characterized	3,25	0,00	4,01	2,90E-02
XCC2091		tRNA-Leu_NC_003902.1	Not analyzed	Poorly characterized	11,78	0,58	3,95	1,59E-03
XCC3410	*nagA*	N-acetylglucosamine-6-phosphate deacetylase	G	Energy, carbon and central intermediary metabolism	4,36	0,09	3,93	1,09E-02
XCC1792		tRNA-Arg_NC_003902.1	Not analyzed	Poorly characterized	9,22	0,41	3,80	5,89E-04
XCC2265		pectinesterase	G	Pathogenicity and adaptation	19,80	1,36	3,78	1,87E-06
XCC0879		tRNA-Thr_NC_003902.1	Not analyzed	Poorly characterized	3,92	0,09	3,72	3,06E-02
XCC2535		hypothetical protein	Unclassified	Poorly characterized	5,06	0,23	3,65	9,80E-03
XCC3316	*btuB*	TonB-dependent receptor	P	Transport	50,91	4,01	3,62	1,35E-13
XCC0308		dehydrogenase	I	Energy, carbon and central intermediary metabolism	4,61	0,33	3,43	2,16E-02

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC0234		hypothetical protein	R	Poorly characterized	5,32	0,47	3,14	4,14E-02
XCC2884	*bioA*	adenosylmethionine--8-amino-7-oxononanoate aminotransferase BioA	H	Biosynthesis of amino acids, cofactors, prosthetic groups and carriers	16,34	1,64	3,14	6,83E-05
XCC3183		hypothetical protein	O	Protein degradation and post-translational modifications	14,33	1,49	3,01	5,89E-04
XCC2667		serine protease	O	Protein degradation and post-translational modifications	6,24	0,64	2,96	1,24E-02
XCC2608		hypothetical protein	S	Poorly characterized	12,87	1,65	2,93	5,14E-04
XCC1816		hypothetical protein	R	Poorly characterized	21,54	2,73	2,81	1,53E-03
XCC0253		hypothetical protein	M	Cell envelope and structure	7,71	0,94	2,76	4,36E-02
XCC3283	*kch*	ion transporter	P	Transport	7,07	0,87	2,74	1,26E-02
XCC1341		hypothetical protein	S	Poorly characterized	7,33	0,97	2,69	1,82E-02
XCC2136	*zwf*	glucose-6-phosphate 1-dehydrogenase	G	Energy, carbon and central intermediary metabolism	7,10	1,19	2,66	2,61E-02
XCC4119	*xypA*	hexuronate transporter	GEPR	Pathogenicity and adaptation	13,67	1,90	2,66	1,47E-03
XCC4224		hypothetical protein	T	Signal transduction and regulation	15,72	2,52	2,64	1,46E-03
XCC0076		metalloprotease	E	Protein degradation and post-translational modifications	6,91	1,17	2,55	2,63E-02
XCC1704	*truD*	tRNA pseudouridine synthase D	S	Transcription and translation	10,66	1,48	2,54	2,15E-02
XCC0113		hypothetical protein	Unclassified	Poorly characterized	8,62	1,34	2,49	1,72E-02
XCC2706	*cls*	cardiolipin synthase	I	Cell envelope and structure	11,61	2,12	2,47	3,72E-03
XCC4024		hypothetical protein	T	Signal transduction and regulation	6,88	1,11	2,47	2,86E-02
XCC3097	*pilB*	pilus biogenesis protein	NU	Pathogenicity and adaptation	7,14	1,24	2,40	3,01E-02
XCC3484		hypothetical protein	Unclassified	Poorly characterized	12,11	2,14	2,32	2,60E-02

Gene ID	Name	Annotation	COG description	Biological role	Mean Expression		log_2FC	Adj.
					L-vir	H-vir	(L-vir/H-vir)	p-value
XCC2671	*mocA*	oxidoreductase	C	Energy, carbon and central intermediary metabolism	12,25	2,47	2,26	1,22E-02
XCC1406	*ygdR*	oligopeptide transporter	E	Transport	12,56	2,70	2,15	9,09E-03
XCC2866		hypothetical protein	Unclassified	Poorly characterized	25,19	5,66	2,12	3,09E-03
XCC1895		hypothetical protein	J	Protein degradation and post-translational modifications	87,05	20,13	2,07	2,34E-04
XCC2407		hypothetical protein	S	Transport	16,81	3,87	2,04	5,33E-03
XCC0004	*gyrB*	DNA gyrase subunit B	L	DNA replication and metabolism	2,38	3,66	1,90	1,10E-04
XCC3625	*aceE*	pyruvate dehydrogenase subunit E1	C	Energy, carbon and central intermediary metabolism	18,70	5,80	1,63	3,91E-02

DEGs involved in pathogenicity and adaptation

Included in this category are 25 DEGs coding for known and candidate type III effectors, as well as other proteins known to play a role in plant-pathogen interactions, such as secretion system individual components, cell-wall degrading enzymes, detoxification enzymes and quorum sensing related proteins (Figure 24). Among this set of genes, 13 were repressed, whereas 12 were induced in *Xcc* L-vir strain.

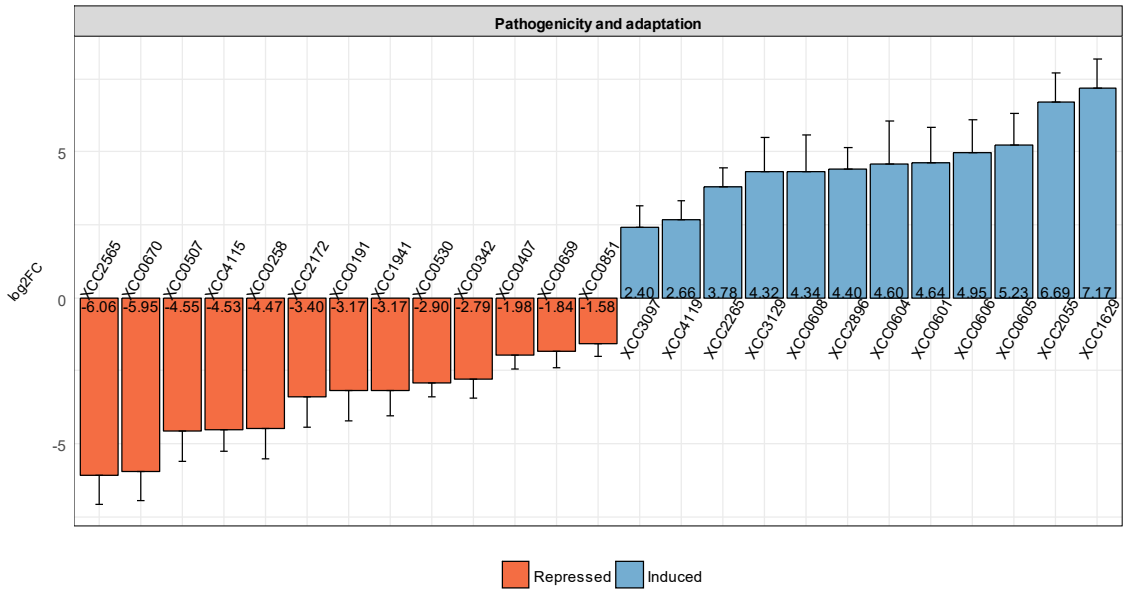


Figure 25 - Differentially expressed genes (DEGs) involved in pathogenicity and adaptation. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

Among the DEGs involved in pathogenicity and adaptation there were five known Type III effectors. *xopE2* (also named *avrXccE1*) and x*opD* were induced in *Xcc* L-vir strain, while *xopAC* (also named *avrAC*)*, xopR* and *xopX* were repressed in this strain.

xopE2 (XCC1629), coding for a putative transglutaminase described as an avirulence protein, was recorded as the most induced DEG in L-vir *Xcc* strain used in this study. XopE2 belongs to the XopE family of T3Es that is highly similar to HopX (AvrPphE) of *Pseudomonas syringae* pv. *tomato* (*Pst*). Members of the widespread HopX (AvrPphE) family are modular proteins composed of a conserved putative cysteine-based catalytic triad and a conserved potential target/cofactor interaction domain (Nimchuk *et al.*, 2007).

XopE2 was first described in *X. euvesicatoria* strain 85-10 as a protein possessing a conserved putative N-myristoylation motif responsible for its targeting to the plant cell membrane that is also present in the effector XopJ, a member of the YopJ/AvrRxv family of acetyltransferases (Thieme *et al.*, 2007).

Using an *Arabidopsis thaliana* protoplast transient expression system, XopE2 from *X. euvesicatoria* significantly reduced the induction of a flg22-inducible reporter gene construct, suggesting that this effector has the ability to suppress PTI (PAMP-Triggered Immunity) (Popov *et al.*, 2016). An attenuated *Pst* strain carrying *xopE2* was associated with a decrease of callose deposition and an increase of disease symptoms in susceptible pepper plants (Popov *et al.*, 2016). However, mutation of this T3E in *X. vesicatoria* did not lead to any significant virulence loss against pepper plants. Interestingly, in *A. thaliana*, this effector did not affect the flg22-induced MAP kinase phosphorylation, indicating that it does not interfere with FLS2 receptor or components of the FLS2 receptor complex, suggesting that it targets other receptors, or that the receptors in *A. thaliana* are not sufficiently similar to those of pepper plants to recognize this effector (Popov *et al.*, 2016).

XopE2 has also been described in *X. axonopodis* pv. *citri* (da Silva *et al.*, 2002) and in the three *X. campestris* pathovars (*X. campestris pv. campestris*, *X. campestris* pv. *raphani* and *X. campestris* pv. *incanae*) (Roux *et al.*, 2015). Contrarily to what has been described for *X. euvesicatoria*, in *Xcc* strain 8004, expression of *xopE2* by naturally virulent strains was found to confer avirulence to the Chinese cabbage cv. Zhongbai-83, thus determining host specificity (He *et al.*, 2007b).

The results obtained in the present study are consistent with the hypothesis of He *et al.* (2007), since a higher production of XopE2 is associated with the lowest level of virulence.

Regulation of *xopE2* expression appears to occur due to the presence of a Plant Inducible Promotor (PIP)-box containing promotor, that has been shown to be bound by the AraC-type transcriptional activator HrpX (Koebnik *et al.*, 2006). Additionally, its expression appears to be regulated by *rsmA* (repressor of secondary metabolism), since *rsmA Xcc* mutants had reduced expression of *xopE2* when compared to the wild-type (Chao *et al.*, 2008).

XopD encoded by locus XCC2896 is a Small Ubiquitin-related Modifier (SUMO) protease that inhibits host transcription to suppress defense hormone signaling. This T3E was first described in *X. euvesicatoria* (*Xcv*) strain 85-10 as a SUMO protease. With a structure resembling that of ubiquitin, SUMO proteins are ubiquitously distributed in eukaryotes and function as a reversible post-translational protein modifier (Geiss-friedlander & Melchior, 2007). Like ubiquitylation, sumoylation results in the formation of an isopeptide bond between the C-terminal Gly residue of the modifier protein and the ε-amino group of a Lys residue in the acceptor protein (Geiss-friedlander & Melchior, 2007).

In addition to its C-terminal SUMO protease domain, XopD has a unique N-terminal region with a nonspecific DNA-binding domain (DBD) that determines host range and a central domain with two EAR motifs (Ethylene-responsive element binding factor-associated Amphiphilic Repression), which

are found in plant repressors that regulate stress-induced transcription, which suggests that XopD might repress host transcription on *Xcv* infected leaves. In fact, XopD was found to repress production of the plant hormone salicylic acid (SA), a plant defense hormone that limits the spread of biotrophic pathogens, and SA-dependent gene expression, thus promoting *Xcv* multiplication and delaying disease progression (Kim *et al.*, 2008). Additionally, XopD has been found to directly target the tomato ethylene responsive transcription factor SIERF4 to suppress ethylene production, which is also required for anti-*Xcv* immunity (Kim *et al.*, 2013).

In *X. campestris* pv. *campestris* strain 8004, XopD was found to have SUMO protease activity targeting HFR1, a basic helix-loop-helix transcription factor involved in light-signaling pathway, in *A. thaliana* (Tan *et al.*, 2015). Using a transgenic approach, based on the construct of plants carrying a vector expressing XopD, those authors recorded a XopD-induced elevation of host defense-response gene expression in a SA-dependent manner (Tan *et al.*, 2015). Similarly, expression of XopD in transgenic plants suppressed the multiplication of the *Xcc* 8004 strain *in planta* at 5 days post inoculation, suggesting a reduction in overall virulence (Tan *et al.*, 2015).

In the present study, expression of *xopD* was found to be induced in the lowest virulence strain, supporting the hypothesis that XopD is associated to reduced virulence, possibly through the activation of plant cell death promoted by ethylene accumulation.

XopAC, a protein containing leucine-rich-repeats (LRR), was found to confer avirulence on *Xcc* strain 8004 upon inoculation of *A. thaliana* ecotype Col-0. Interestingly, avirulence was only recorded after inoculation by piercing of the leaves and not following leaf infiltration, suggesting that this avirulence protein is only recognized by the host in vascular tissue (Xu *et al.*, 2008). More recently, XopAC was found to confer avirulence to both the vascular pathogen *Ralstonia solanacearum* and the mesophyll-colonizing pathogen *Pseudomonas syringae* indicating that XopAC-specified effector-triggered immunity is not specific to the vascular system (Guy *et al.*, 2013). In the same study, both the Receptor-like cytoplasmic kinases (RLCK) interaction domain and the uridylyl transferase domain of XopAC were found to be required for avirulence (Guy *et al.*, 2013). Plasma membrane-localized XopAC interacts with the RLCKs PBL2 (PBS [AvrPphB susceptible]1-like 2) and RIPK (RPM1[Resistance to Pseudomonas Syringae pv. Maculicola 1]-induced protein kinase) (Guy *et al.*, 2013). RLCKs are activated via PRR signaling upon PAMP recognition, further inducing defense gene expression and therefore playing a central role in PAMP-Triggered Immunity in plants (Yamaguchi *et al.*, 2013).

In this study, *xopAC* (XCC2565) was found to be repressed in L-vir. A diminished expression may be a result of a lack of RLCK activation on the host cells, which in turn can be the consequence of poor PAMP recognition. This hypothesis could be supported by the repression of *fliC* (XCC1941, further described below) coding for flagellin in *Xcc* L-vir strain. Although the exact mechanism behind

regulation of *xopAC* has not yet been revealed, expression of this gene appears to be under the control of *hrp* regulon genes, namely *hrpG* and *hrpX* (Xu, *et al*. 2008).

XCC0530, coding for an ortholog of the T3E XopX of *X. euvesicatoria* (*Xcv*), was also repressed in L-vir *Xcc* strain. XopX, originally identified in *Xcv*, was found to contribute to the virulence of *Xcv* on host pepper and tomato plants, suggesting that it targets basic innate immunity in plants, resulting in enhanced plant disease susceptibility (Metz *et al.*, 2005). More recently, XopX was shown to promote ethylene production and plant cell death during *Xcv* infection of susceptible tomato, confirming its contribution for the development of *Xcv* induced disease symptoms. Additionally, XopX was responsible for overturning PTI, through the suppression flg22-induced production of reactive oxygen species (ROS), although it also appeared to promote transcriptional activation of other PTI-associated genes, such as those involved in salicylic acid and ethylene biosynthesis (Stork *et al.*, 2015). This apparent contradiction is consistent with a model that states the redundancy of the hosts defense mechanism pathways, in which the blockage of a given defense pathway by the pathogens' effectors leads to the activation of a different pathway (Sato *et al.*, 2010). In *X. oryzae* pv. *oryzae*, XopX was also found to suppress the rice defense responses induced by the activity of the bacterial cell wall degrading enzyme LipA, such as callose deposition and programmed cell death (Sinha *et al.*, 2013). Although it has not been identified, the molecular target of XopX appears to be involved in PTI responses. The reduced expression of *xopX* by *Xcc* L-vir strain used in this study could ultimately result in a reduced capacity to impair the hosts' inate immunity responses.

XCC0258, coding for a protein ortholog of XopR found in *X. oryzae* pv. *oryzae* (*Xoo*), was repressed in *Xcc* L-vir strain. In *Xoo*, XopR has been shown to suppress PTI responses in rice through its association with receptor-like cytoplasmic kinase BIK1 (Wang *et al.*, 2016). BIK1 is a component of the FLS2 immune receptor complex that directly phosphorylates the NADPH oxidase RbohD at specific sites, controlling ROS generation, causing active reduction of stomatal aperture to prevent the entry of microbes (Wang *et al.*, 2016).

Taken together, these findings suggest that the low virulence phenotype observed in *Xcc* L-vir is the combined result of the overexpression of avirulence proteins (XopE2 and XopD) associated to the repression of virulence proteins targeting PTI host responses (XopAC, XopX and XopR). Additionally, the results obtained in this work confirm the function of these T3Es as modulators of PTI, instead of ETI (Effector-Triggered Immunity).

In addition to the known T3 effectors, other pathogenicity-related and adaptation-related proteins were also identified as differentially expressed in the *Xcc* strains studied.

The flagellin coding gene *fliC* (XCC1941) was repressed in L-vir. Flagellin is the building block of flagella, crucial for their structure and function. Taken together, these results suggest a diminished motility of *Xcc* L-vir strain, associated to a less recognizable phenotype, which could have serious impact on pathogenicity and disease progression.

The virulence protein VirB6, coded by *virB6* (XCC3129) was also found to be induced in L-vir. VirB6 is a key component of Type IV secretion systems (T4SS). T4SSs are generally made up of a core set of 12 proteins, named VirB1–VirB11 plus VirD4 with the following structural organization: (i) a set of three cytoplasmic ATPases (VirB4, VirB11 and VirD4) that energize secretion; (ii) a periplasmic core complex made up of 14 repeats of a VirB7–VirB9–VirB10 trimer, in which VirB10 inserts into both inner and outer membranes while VirB7 is an outer membrane lipoprotein; (iii) an inner membrane complex that includes VirB3, VirB6 and VirB8; (iv) an extracellular pilus formed by VirB2 and VirB5; and (v) the periplasmic transglycosylase VirB1 (Souza *et al.*, 2015).

In *Xcc* strain 8004, deletion of T4SS *virB* cluster and *virD4* did not affect virulence against several cultivars of the susceptible hosts *Brassica oleracea*, *B. rapa* and *Raphanus sativus*, suggesting that the T4SS is not directly involved in virulence (He *et al.*, 2007b). Despite the involvement of T4SS in the infection process of many bacterial pathogens has been established (like *Agrobacterium tumefaciens, Helicobacter pylori and Legionella pneumophila*), a recent *in planta* analysis of *X. axonopodis* subsp. *citri* virB cluster expression failed to correlate T4SS with pathogenicity (Jacob *et al.*, 2014). The results obtained in this study showed the overexpression of a T4SS component in the least virulent *Xcc* strain, thus contributing to the growing idea that T4SS is not directly involved in pathogenicity in *Xanthomonas* spp.

xpsD (XCC0670), coding for a general secretion pathway protein D, was found to be repressed in L-vir. The *xps* gene cluster is one of the two Type II secretion systems found in *Xcc*. The type II secretion system (T2SS) is the main terminal branch of the general secretory pathway in proteobacteria, mediating the transport of proteins into the extracellular space following their N-terminal signal peptide–dependent deposition into the periplasm. T2SS are thus involved in pathogenicity, by exporting toxins, proteases, lipases, and phospholipases, as well host cell wall–degrading enzymes (CWDEs) such as cellulases, pectinases and xylanases (Lu *et al.*, 2008). Variation in four residue positions of *xpsD* have also been related to host tissue-specificity (Lu *et al.*, 2008). *xpsD* was found slightly induced when *Xcc* strain ATCC33913^T was grown in the presence of cabbage xylem sap (de Bernonville *et al.*, 2014).

pilB (XCC3097) coding for pilus biogenesis protein was found to be induced in L-vir strain. PilB is an ATPase component of the bacterial type IV pilus (T4P), a molecular nanomachine also related to the ubiquitous T2SS (Dunger *et al.*, 2016). This protein is responsible for catalyzing the incorporation

of pilin subunits during pilus polymerization/extension and its regulation appears to be driven by a complex set of protein-protein interactions modulated by the secondary messenger c-di-GMP and/or other metabolic or environmental signals (Dunger *et al.*, 2016). In *Xcc* 8004, PilB is thought to interact with two proteins containing a GGDEF domain encoded by loci XC_0249 and XC_0420 (Dunger *et al.*, 2016). GGDEF domains (named after the conserved central sequence pattern GG[DE][DE]F) are responsible for the synthesis of cyclic di-GMP biosynthesis from two GTP molecules and are usually found in proteins involved in cyclic di-GMP turnover (Sondermann *et al.*, 2012). The ortholog of XC_0420 protein in *Xcc* ATCC33913^T is encoded by locus XCC0407. A mutant derived from *Xcc* strain 8004 lacking this GGDEF containing protein (*Xcc*8004ΔXC_0420) was found to have significantly reduced virulence against the susceptible host *R. sativus*, suggesting that it is required for full virulence (Ryan *et al.*, 2007). Consistent with this hypothesis is the fact that this locus was differentially expressed among the two *Xcc* strains used in this study, being repressed in L-vir strain. Additionally, the fact that both PilB and XCC0407 were differentially expressed in strains in contrasting patterns may suggest that this protein-protein interaction may be regulated by a negative feedback mechanism, most via cyclic-di-GMP turnover.

Cyclic-di-GMP is sensed by transcriptional factors, PilZ domain-containing proteins, GGDEF and/or EAL (named after its conserved residues) domain-containing proteins, or riboswitches, which then interact with a downstream target affecting biofilm formation, motility, virulence and cell cycle (Sondermann *et al.*, 2012). This secondary messenger is known to play a central role in the regulation of *hrp* gene expression and host cell wall degrading enzyme (CWDE) biosynthesis via *clp* regulation (He *et al.*, 2007a) .

CWDEs, usually secreted by T2SS, are known to play a central role in pathogenesis, breaking down the network of cellulose, non-cellulosic polysaccharides, proteins, and aromatic substances existing at the plant cell wall (Jacques *et al.*, 2016).

In *Xcc* L-vir, XCC0659 and XCC0851 (*prtA*), both coding for extracellular serine proteases of the subtilisin family, were found to be down-regulated in *Xcc* L-vir. These proteins are part of the quorum sensing pathways described for *Xcc*, suggesting that they are involved in regulatory mechanisms depending on the bacterial population density. XCC0659 was reported to be repressed when *Xcc* ATCC33913^T was grown in the presence of cabbage xylem sap (de Bernonville *et al.*, 2014). XCC0659 orthologous protein CylA (Cytolysin activator) of *Enterococcus faecalis* has been found to function as a bacterial toxin and is often screened for as a virulence gene (Coburn & Gilmore, 2003). In previous studies, XCC0851 was slightly induced the presence of cabbage xylem sap (for *Xcc* strain ATCC33913^T) as well as in the minimal medium MMX (for *Xcc* strain 8004), suggesting a transcriptional activation of this locus associated with pathogenesis and adaptation (Liu *et al.*, 2013; de Bernonville *et al.*, 2014).

XCC2265, coding for a pectinesterase, was induced in *Xcc* L-vir strain. The promoter of the gene coding for the ortholog protein XC_1850 in *Xcc* strain 8004, was shown to include a PIP-box, a plant-inducible-promotor sequence known to be activate during pathogenesis. This gene was found to be slightly repressed when the strain was grown in MMX, a culture medium that induces T3Es (Liu *et al.*, 2013).

xynA (*xyn10C*) (XCC4115) coding for an endo-1,4-beta-xylanase A (ring canal kelch-like protein) was repressed in L-vir *Xcc* strain. This gene is part of the *xytB* operon, one of the xylan degrading operons found in *Xcc* (Déjean *et al.*, 2013). XCC4119, coding for a Major Facilitator Superfamily (MFS) transporter, highly similar to *exuT* (hexuronate transporter of *Ralstonia solanacearum*), recently named XypA, was found to be induced in L-vir *Xcc* strain. Also a component of the xylan degrading operon *xytB*s, XypA is thought to be involved in the transportation of glucuronic acid resulting from xylan degradation (Déjean *et al.*, 2013). Induction of xylan degrading operons usually occurs in the presence of xylo-oligosaccharides, while it is repressed by XylR LacI-type repressor (Déjean *et al.*, 2013).

Low expression of some CWDE could be explained by the recorded repression of *xpsD*, a key component of T2SS playing a critical role in the export of CWDE (Jacques *et al.*, 2016).

Lipopolysaccharides (LPSs) are a key structural component of the bacterial cell wall, highly conserved in prokaryotes and important PAMPs, triggering the hosts' primary immunity upon recognition (Jones & Dangl, 2006). The over-expression of genes involved in LPS biosynthesis, leading to an over-production of this PAMP, could ultimately lead to a faster recognition from the host and a quicker PTI response.

In *Xcc* L-vir, *gmd* (XCC0606), coding for a GDP-mannose 4%2C6-dehydratase, *wzt* (XCC0601) coding for an ABC transporter (ATP-binding cassette transporter) and *rmd* (XCC0608), coding for a UDP-glucose 4-epimerase were found to be induced. Additionally, the poorly annotated genes *wxcE* (XCC0605) and *wxcD* (XCC0604), were also up-regulated in L-vir. These five genes belong to the *wxc* gene cluster, responsible for lipopolysaccharide (LPS) biosynthesis in *Xcc*. The *wxc* gene cluster comprises 15 genes, divided into 3 main regions, on the basis of mutant LPS phenotypes (Vorhölter *et al.*, 2001). Genes from region 1 are necessary for the biosynthesis of the water-soluble LPS O-antigen, genes from region 2 direct the synthesis of the sugar GDP-D-rhamnose and genes from region 3 have similarities to genes that code for enzymes that modify nucleotide sugars. The genes induced in L-vir strain are part of *wxc* regions 1 and 2, suggesting that they are involved in the synthesis of both O-antigen and GDP-D-rhamnose.

In addition to the fact that LPS serves as PAMP, changes in LPS composition have been associated to tissue specialization and host jumps in *X. oryzae* and *X. vasicola*, respectively, determining host specificity (Jacques *et al.*, 2016). Although no evidence of LPS-driven host specificity has been found in *X. campestris*, the fact that it is plausible in other xanthomonads makes LPS biosynthesis gene clusters ideal candidates for survey in closely related xanthomonads with differing host range.

Contrastingly, XCC0342, coding for an uncharacterized xanthomonadin biosynthesis related protein, was repressed in *Xcc* L-vir strain. Although useful as a chemotaxonomic and diagnostic marker, this membrane-bound, halogenated, aryl-polyene pigment is not essential for *in planta* growth of bacterial cells. Several studies point to the idea that xanthomonadins protect bacteria against photooxidative damage, potentially aiding their survival on plant leaf surfaces (Goel *et al.*, 2002)

cpo (XCC2172), coding for a non-heme chloroperoxidase, and *sodC2* (XCC0191), coding for a superoxide dismutase (Cu-Zn family), were repressed in L-vir. The proteins coded by these genes are involved in detoxification mechanisms, providing important defense against the hosts' primary defenses, which often include the production of reactive oxygen species (Jones & Dangl, 2006). The repression could result in a weakened detoxification capacity contributing to a diminished survival rate of the bacterial population in this strain.

Also involved in adaptation mechanisms to atypical conditions, *ndvB* (XCC2055) gene, coding for a cyclic β-(1,2)-glucan synthase was also found to be induced in L-vir strain. In a previous study, disruption of the *Xcc* nodule development B gene (*ndvB*) generated a mutant that failed to synthesize extracellular cyclic b-(1,2)-glucan and was compromised in virulence in the model plants *Arabidopsis thaliana* and *Nicotiana benthamiana* (Rigano *et al.*, 2007). In the same study, cyclic glucan was proven to be systemically transported throughout the plant and to suppress host defenses (Rigano *et al.*, 2007).

XCC0507, coding for an ice nucleation protein, was found repressed in *Xcc* L-vir strain. Ice nucleation proteins are synthetized by several bacterial species and are thought to confer them the ability to order water molecules into an ice-like structure that raises the temperature at which ice nucleates, causing water to freeze at high sub-zero temperatures (Pfeilmeier *et al.*, 2016). Since these bacteria are pervasive on the surfaces of frost-sensitive plants, frost damage caused by induced ice nucleation allows pathogen entry into plant tissue (Gurian-Sherman & Lindow, 1993).

DEGs involved in signal transduction and regulatory functions

Corresponding to 8% of the global DEG set, a total of 13 genes coding for proteins involved in signal transduction and regulatory functions were identified. Of those, seven were repressed while six were induced in *Xcc* L-vir strain, as depicted in Figure 25.

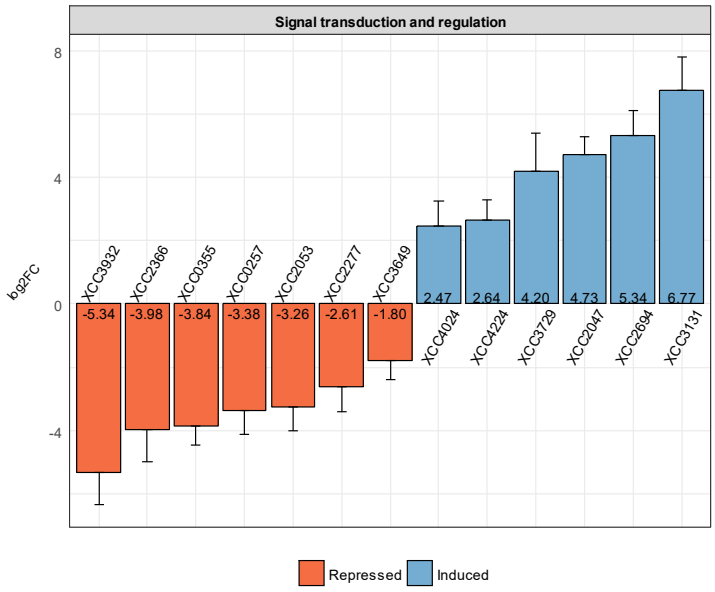


Figure 26 - Differentially expressed genes (DEGs) involved in signal transduction and regulation. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each gene) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

In bacteria, two-component signal transduction systems (TCSs) are the central sensory-responsive mechanisms, being involved in several physiological pathways (Qian *et al.*, 2008). Typically, these systems comprise a membrane bound sensory histidine kinase (HK) and a cytoplasmic response regulator (RR) (Stock *et al.*, 2000). Upon specific stimulation, the HK can be auto-phosphorylated and transfer the phosphoryl group to the cognate RR. The output domain of the RR modulates the expression of downstream genes or cellular machinery, causing adaptive changes (Stock *et al.*, 2000). Along with other signaling proteins, the total number of TCSs encoded by a bacterial genome can thus be used as a rough estimate of the adaptive potential of the organism (Galperin, 2005). In *Xanthomonas* spp. genomes, the predicted number of TCSs varies between 92 and 121 (Qian *et al.*, 2008). In the genome of *Xcc* 8004, a total of 106 putative TCS genes have been predicted (Qian *et al.*, 2005).

creC (XCC2694), coding for a two-component sensory histidine kinase, was found to be highly up-regulated in L-vir. This sensory histidine kinase, along with its corresponding response regulator CreB form a two-component signal transduction system, highly conserved among many gram-negative bacteria. Despite this conservation, the mechanisms regulating the activation of *creBC* vary significantly between organisms. Overall, *creBC* responds to metabolic signals or peptidoglycan stress

(Huang *et al.*, 2015). Peptidoglycan stress can result from host-mediated peptidoglycan degradation, through hydrolytic activities that facilitate decomposition of bacterial matrices and generation of soluble PAMPs that trigger plant immunity (Liu *et al.*, 2014). Over-expression of *creC* in the *Xcc* L-vir strain was accompanied by *creD* (XCC2692) repression. As further discussed on the section about DEGs involved in cell envelope and structure, these results are in accordance to what has been described for *S. maltophilia* but distinct from *P. aeruginosa*, suggesting that the CreBC TCS function may be distinct among different families of gram-negative bacteria (Huang *et al.*, 2015).

Another TCS, involved in bacterial chemotaxis and motility, is the system CheA-CheYBV, which relies on methyl-accepting chemotaxis proteins (MCPs) as the primary receptors of the chemotaxis signal transduction cascade that process environmental and intracellular sensory (input) signals (Wadhams & Armitage, 2004). An adaptor protein, CheW, helps to link the MCPs to the cytoplasmic HK, CheA, and two RRs compete for binding to CheA. One RR is a single-domain, flagellar-motor-binding protein, CheY. Phosphorylated CheY binds the switch protein FliM on the flagellar motor and causes a reversal in the direction of motor rotation. This reversal causes tumbling and lets the bacteria re-direct their movement towards a more favorable environment. While *tsr* (XCC2047) coding for an MCP was induced, response regulator *cheY* (XCC2053) was repressed in L-vir. This expression pattern suggests that although the sensory domain of the TCSs is functioning properly, the diminished expression of the response regulator could affect the ability of the bacteria to reverse the direction of flagellar motor, hindering bacterial motility. *cheY* and *fliC* have also been reported to be repressed when *Xcc* strain 8004 was grown in the nutritionally scarce MMX medium when compared to a rich medium, thus a transcriptional regulation of these genes appears to be the result of a nutritionally challenged ecological niche (Liu *et al.*, 2013).

An additional putative sensor kinase, encoded by XCC3131, was also found highly induced in the L-vir *Xcc* strain. This protein contains a typical MASE1 domain (Membrane Associated Sensor), a predicted integral membrane sensory domain found in histidine kinases, diguanylate cyclases and other bacterial signaling proteins, predicted to ligate aromatic compounds (Galperin, 2004). While an ortholog of this gene has been found in *X. vesicatoria* (XCV3392), no further information on function was available.

Signaling via secondary messenger cyclic di-GMP has been shown to regulate biofilm formation, motility, virulence, the cell cycle, differentiation, and other processes and plays an important role in *Xcc* pathogenesis (Ryan *et al.*, 2007; Römling *et al.*, 2013). Cellular levels of cyclic di-GMP are controlled through synthesis, catalyzed by the GGDEF protein domain, and degradation by EAL or HD-GYP domains (Ryan *et al.*, 2007).

In *Xcc* L-vir, two GGDEF containing hypothetical proteins were induced – XCC4224 and XCC3729. In addition to the GGDEF domains, these proteins also include sensory domains (such as PAS fold or dCACHE domains), suggesting that they play an important role in the perception of stimuli that may trigger the synthesis of cyclic di-GMP. Elevated cyclic di-GMP levels or increased GGDEF protein expression is frequently associated with the onset of reduced motility and biofilm formation in several bacterial species (Tan *et al.*, 2014). Conversely, phosphodiesterase-dependent reduction of cyclic di-GMP levels by EAL- and HD-GYP-domain proteins is often related to increased motility and virulence.

Activation of signal transduction systems often results in the binding of a small molecule to a transcriptional regulator to modulate its activity on gene expression. Interestingly, most transcriptional regulators qualified as DEGs were repressed in the lowest virulence strain – *pobR*, XCC3932, XCC2366, *suhB*, *lrp* and XCC0257.

pobR (XCC0355) and XCC3932 are AraC-family transcriptional activators. This family of transcriptional regulators is characterized by the presence of a DNA binding domain with significant amino acid sequence homology extending over a 100-residue stretch (Gallegos *et al.*, 1997). In *X. campestris*, the AraC-like regulator HrpX is activated by the response regulator HrpG, part of the HrpG/HrpB two-component signal transduction system (Mole *et al.*, 2007). Although the mechanisms behind this regulation have not been fully clarified, AraC -like regulator HrpX is responsible for the activation of T3SS genes, thus playing a central role in pathogenicity (Büttner & Bonas, 2010)

In *Xcc* ATCC33913^{T}, PobR has been shown to play a key regulatory role in the degradation of the phenolic compound 4-hydroxybenzoate (4-HBA) by PobA, encoded by *pobA* (XCC0356) (Wang *et al.*, 2015). Deletion of *pobR* significantly reduced the expression of *pobA*, whereas overexpression of *pobR* resulted in the upregulation of *pobA* (Wang *et al.*, 2015). In addition to being actively produced by *Xcc* (to serve as an intermediate in energy production metabolic pathways, such as the TCA cycle or aerobic respiration), there is also an uptake of 4-HBA by *Xcc* in certain conditions, such as the presence of plant hydrolysates (Wang *et al.*, 2015). Plant phenolic compounds are essential for reproduction and growth, as well as defense against pathogens, since many pathogens are susceptible to aromatic stress (Wang *et al.*, 2015). On the other hand, the AraC transcriptional regulator encoded by XCC3932 has not been linked to any biological process so far.

XCC2366, coding for a putative TetR/AcrR family transcriptional regulator, was also repressed in the lowest virulence strain. Although this family of regulators is involved in the regulation of an extensive range of cellular processes, the most frequent function of its members is the regulation of efflux pumps and transporters involved in antibiotic resistance and tolerance to toxic chemical compounds (Deng *et al.*, 2013)

suhB (XCC2277) codes for an extragenic suppressor protein, that is repressed in L-vir. These proteins are responsible for the annulment of mutations. This protein is also part of the metabolic pathways responsible for inositol phosphate metabolism and streptomycin biosynthesis. There was no information available concerning the role of *suhB* in *Xcc.*

lrp (XCC3649) codes for a leucine responsive regulatory protein, a member of Lrp/AsnC family of transcriptional regulators, widely distributed among prokaryotes (Thaw *et al.*, 2006). Lrp is known to affect expression of many operons and genes in *E. coli*, including the repression of transcription of the operon *ilvGMEDA* about three-fold, by binding two sites in the DNA region between the attenuator and the *ilvG* gene (F.Wendisch, 2007). The results obtained in this study contrast to what has been described for *E. coli*, since the repression of *lrp* in L-vir was also accompanied by a repression of *ilvG* and ilvE, suggesting a complex mechanism for the regulation of *ilvGMEDA* operon. The gene sequence of *lrp* has been used for the establishment of genetic relationships among *Xanthomonas* spp., since it was considered a housekeeping gene (Cubero & Graham, 2004). The results obtained in this study show that *lrp* expression can vary among different strains.

XCC0257, coding for a nitrile hydratase activator, was also repressed. Nitrile hydratase (NHase) activators are essential for the functional expression of NHase, an enzyme that catalyzes the hydration of nitriles to the corresponding amides (Lu *et al.*, 2003). The homolog of XCC0257 in *Xcc* strain 8004 (encoded by locus XC_0267) was described as a putative Zn^{2+} uptake-related gene similar to YciC, a component of the low-affinity Zn^{2+}-uptake system YciABC of *B. subtilis* (Huang *et al.*, 2008).

XCC4024, that is induced in L-vir, codes for a hypothetical protein containing a Universal stress protein (Usp) family domain. Usp was discovered in *E. coli*, as a protein whose expression is enhanced when the cell is exposed to several stress agents or conditions, such as stationary phase, starvation for carbon, nitrogen, phosphate, sulphate and amino acids, and exposure to heat, oxidants, metals, uncouplers, polymyxin, cycloserine, ethanol, antibiotics and other stimulants (Kvint *et al.*, 2003). Although it is clear that UspA enhances the rate of cell survival during prolonged exposure to such conditions, the mechanisms behind its stress endurance activity remain unclear (Nyström & Neidhardt, 1994).

DEGs involved in transport

Of the 13 DEGs identified involved in transport processes, corresponding to an additional 8% of the global DEG set, nine were repressed in *Xcc* L-vir strain, while the remaining five were induced in this strain (Figure 26).

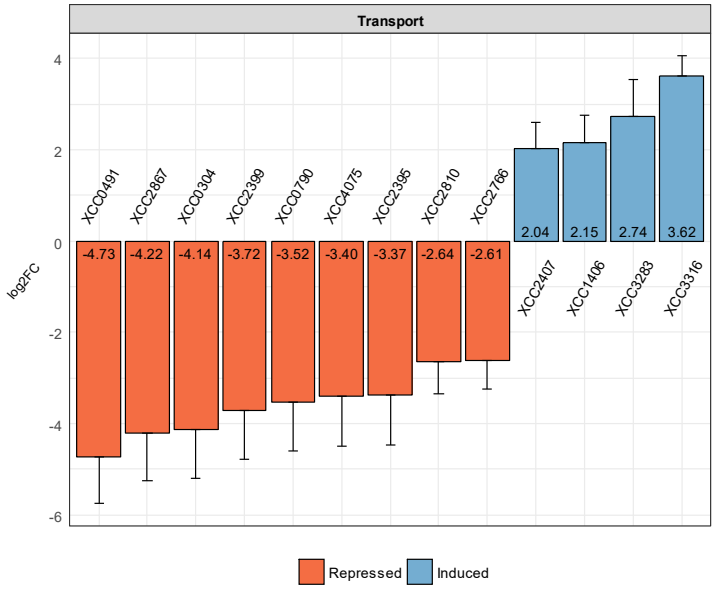


Figure 27 - Differentially expressed genes (DEGs) involved in transport. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

XCC0491 codes for a hypothetical protein that can also be involved in amino acid biosynthesis and transport, according to COG description. This protein contains a transmembrane DUF2752 domain (DUF – Domain of Unknow Function), suggesting that it is a membrane protein. This gene was also found to be repressed when *Xcc* strain ATCC33913^T was grown in the presence of cabbage xylem sap (de Bernonville *et al.*, 2014).

Of the DEGs coding for Inorganic ion transporters, three were repressed (XCC0790, XCC4075 and XCC2810) and only one (XCC3283) was induced in *Xcc* L-vir strain.

XCC0790 codes for ApaG protein. This protein belongs to a superfamily that has been described in several Gram-negative bacteria, including *Escheria coli*, *Salmonella* spp., *Pseudomonas* spp., *Yersinia* spp., *Erwinia* spp., *Vibrio* spp. and *Xanthomonas* spp. (Cicero *et al.*, 2007). In *Salmonella typhimurium*, mutagenesis of the ApaG homolog reduced Co^{2+} tolerance and reduced Mg^{2+} efflux (Gibson *et al.*, 1991). In *X. axonopodis* pv. *citri*, studies on the structure of ApaG suggested that it is involved in the mediation of protein-protein interactions (Cicero *et al.*, 2007).

XCC4075 codes for a putative manganese efflux pump MntP. MntP, also called YebN, was described in *X. oryzae* pv. oryzae as a novel Mn^{2+} export system, playing an important role in Mn^{2+} homoeostasis, detoxification of ROS and protection against hypo-osmotic shock (Li *et al.*, 2011). Additionally, YebN mutants showed substantial reduction of growth rate and virulence in rice, suggesting this transporter may be involved in *X. oryzae* pv. *oryzae* fitness in host (Li *et al.*, 2011).

The Mg^{2+} transporter MgtE*,* coded by *mgtE* (XCC2810), was also repressed. In addition to regulating the influx of Mg2+ , MgtE has also been reported to downregulate T3SS gene expression in *P. aeruginosa*, decreasing cytotoxicity toward epithelial lung cells (Anderson *et al.*, 2010).

The voltage gated K^+ channel coded by *kch* (XCC3283) was the only transporter found to be induced in the lowest virulence *Xcc* strain. In *E. coli*, mutation of *khc* has been proven to confer susceptibility to millimolar concentrations of K^+, leading initiation and continuation of growth to a halt (Kuo *et al.*, 2003).

Similarly, the oligopeptide transporter *ygdR* coded by XCC1406 was induced in L-vir. *ygdR* codes for a transporter belonging to the Proton-dependent Oligopeptide Transporter (POT) family. Based on their crystal structure, this family of transporters has also been described as part of the Major Facilitator Superfamily (MFS) (Solcan *et al.*, 2012). The over-expression POT family members have been considered a consequence of host-induced metabolism during the infection process of several pathogens including *E. coli* (Weitz *et al.*, 2007), *Campilobacter jejuni* (Vorwerk *et al.*, 2014) and *Francisella tularensis* (Meibom & Charbit, 2010).

Contrarily, XCC2766, named *ynhE* or *sufB*, coding for a cysteine desulfurase, was found to be repressed in L-vir strain. This ABC transporter permease is involved in [Fe-S] cluster assembly, belonging to the SufBCDS operon. [Fe–S] cluster is an essential cofactor of several proteins involved in a wide variety of cellular processes, including electron transfer, gene regulation, DNA synthesis and repair, photosynthesis, nitrogen fixation, and cellular defense against oxidative stress (Fuangthong *et al.*, 2015). In *Xcc*, relative expression of SufB protein was elevated under oxidative stress conditions, such as high concentrations of hydrogen peroxide, while repression of SufBCDS proteins has been linked to the activity of the transcriptional repressor protein IscR (Iron-sulfur cluster regulator), located upstream of the operon (Fuangthong *et al.*, 2015).

TonB-dependent receptors (TBDRs) are outer membrane proteins responsible for the uptake of substrates across the outer membrane into the periplasm. This process requires energy in the form of proton-motive force, and a complex of three inner membrane proteins, TonB-ExbB-ExbD, to transduce this energy to the outer membrane (Noinaj *et al.*, 2010). TBDR proteins consist of two domains, with a C-terminal membrane embedded b-barrel domain that is sealed by the N-terminal plug domain which is responsible for binding the substrate at the extracellular side of the membrane and to interact with TonB-ExbB-ExbD at the periplasmic side of the outer membrane (Noinaj *et al.*, 2010). Initially described as essential for the uptake of ferric chelates and vitamin B_{12}, TBDR are also involved in the transport of nickel complexes, and carbohydrates, into the bacterial periplasm (Schauer *et al.*, 2008). In most cases, the expression of the genes encoding these receptors is under the control

of the Fur (Ferric uptake regulator) repressor and activated under conditions of iron starvation (Noinaj *et al.*, 2010).

TBDRs are overrepresented in *Xanthomonas* spp., as well as in other complex carbohydrates scavenging bacteria, and are often part of carbohydrate utilization loci (CUT – carbohydrate utilization containing TBDR) (Blanvillain *et al.*, 2007). CUT loci are defined by the presence of genes coding for carbohydrate degradative enzymes, inner membrane transporters and sugar related regulators, beside TBDR genes (Blanvillain *et al.*, 2007). In *Xcc* ATCC33913^T genome, 72 predicted TBDR genes were identified, distributed among 59 CUT loci.

In this study, 3 TBDR and 1 pseudo-TBDR were identified as differentially expressed between the two contrasting aggressive strains. Additionally, 24 genes described as part of putative CUT loci were also identified. Among these genes were carbohydrate degradative enzymes, a helicase, a transporter, a virulence protein and several hypothetical proteins. Of the 24 CUT-associated DEGs, 12 were induced in L-vir, while the remaining 16 were repressed.

Of the ten existing copies of the *btuB* gene, encoding for a TBDR responsible for the uptake of vitamin B12, three were identified as DEGs – XCC3316, XCC2867 and XCC2395. While XCC3316 was found induced in L-vir, XCC2867 and XCC2395 were repressed in this strain. XCC3316 was also found to be slightly induced when *Xcc* strain 8004 was grown in MMX medium, suggesting that its activation may correspond to a adaption to a nutritionally scarce environment (Liu *et al.*, 2013). On the other hand, XCC2867 was found induced both during *in vitro* growth in the presence of cabbage xylem sap and *in planta* after inoculation (de Bernonville *et al.*, 2014). XCC2395 also corresponds to the main TBDR in the predicted CUT loci that spans genes XCC2391-XCC2430. In this region, two other genes were found to be differentially expressed – XCC2399 and XCC2407. While XCC2399, coding for a hypothetical protein was repressed in L-vir; XCC2407 also coding an unknown protein was induced in this strain. The pseudo-TBDR XCC0304, lacking the plug domain, was also repressed in L-vir *Xcc* strain. This gene is part of a putative CUT loci comprising locus XCC0301 to XCC0308.

DEGs coding for mobile genetic elements

Fourteen DEGs coding for proteins with functions related to mobile genetic were identified, representing 9% of the total DEG set. Of these genes, 10 were repressed while the remaining four were induced in *Xcc* L-vir strain (Figure 27).

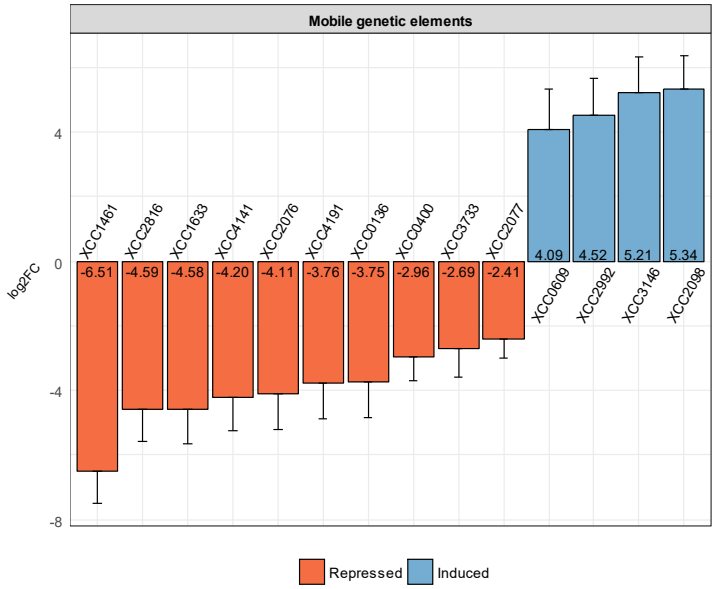


Figure 28 – Differentially expressed genes (DEGs) corresponding or related to mobile genetic elements. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

In L-vir, phage-related protein coding genes *repA* (XCC1461), XCC2076 and *cll* (XCC4191) were found to be repressed, while *intS* (XCC3146) and XCC2992 (T phage-related protein) were induced.

trbP (XCC2077), coding for the conjugal transfer protein TrbP, was repressed in L-vir.

Of the remaining eight DEGs, XCC0609 (IS1479 transposase) and XCC2098 (ISxcd1 transposase) were induced in L-vir, whereas XCC0400, XCC0136, XCC2816, XCC3733, XCC1633 and XCC4141 (coding for IS1404, IS1480 and ISxac3 transposases) were repressed in L-vir.

The high number of mobile genetic elements is usually indicative of a highly flexible genome, which is advantageous for pathogen evolution driven by the need for continuous adaptation to the host in order to evade or suppress coevolving host defense mechanisms (Thieme *et al.*, 2005). Transposases are often part of pathogenicity islands (PAIs), regions of DNA that contain virulence genes that are present in pathogenic strains but absent from non-pathogenic variants of the same or closely related strain (Arnold *et al.*, 2003). In addition to virulence genes and transposases, PAIs also carry other mobile genetic elements, such as insertion sequences and integrases (Arnold *et al.*, 2003). PAIs play an important role in bacterial genome evolution, particularly when it comes to the diversification of the effector repertoire (Arnold & Jackson, 2011). Although the correlation between PAI expression and virulence in Xcc is unclear, the normal expression of PAI SPI-2 (Salmonella Pathogenicity Island 2) was found to be determinant for *Salmonella enterica* virulence inside macrophages of mammalian hosts (Choi *et al.*, 2010).

Differentially expressed genes involved in DNA replication and metabolism

A total of 12 DEGs code for proteins involved in DNA replication and metabolism (Figure 28), corresponding to 7% of the total DEG set.

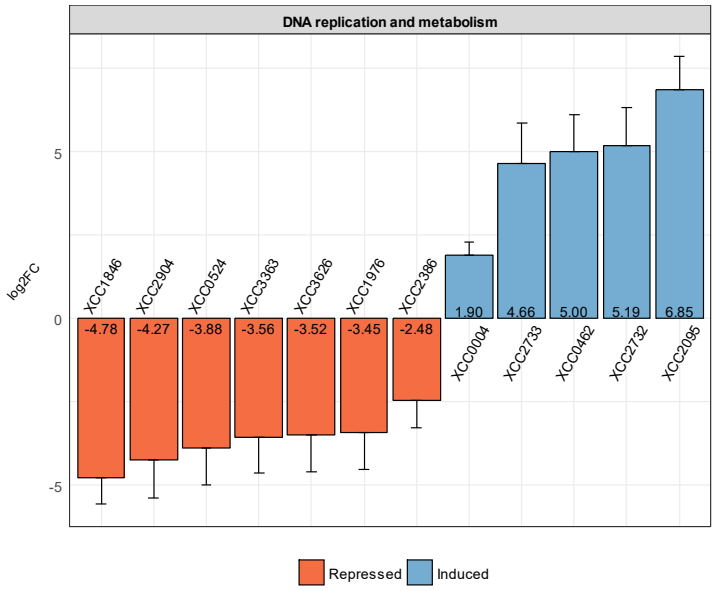


Figure 29 - Differentially expressed genes (DEGs) involved in DNA replication and metabolism. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

Of the DEGs involved in DNA replication and repair, XCC2095, coding for a DNA helicase-like protein, was induced. Contrarily, XCC3626, coding for a RNA-directed DNA polymerase, XCC0524, coding for a helicase, and XCC1976, coding for the recombination factor protein RarA (also named YacJ), were down-regulated in L-vir.

Surprisingly, *gyrB* (XCC0004), the housekeeping gene coding for DNA gyrase subunit B was slightly up-regulated in L-vir strain.

upp (XCC2386), coding for UMP phosphotransferase was repressed in L-vir strain. *upp* has been described in *Xcc* as part of a putative partial CUT locus containing *cirA* (XCC2385), a pseudo-TBDR (Blanvillain *et al.*, 2007).

Several genes involved in bacterial defense mechanisms were also found to be differentially expressed between both *Xcc* strains studied. Type I restriction-modification systems are oligomeric enzymes composed of three different subunits encoded by three closely linked genes – *hsdS*, *hsdM* and *hsdR*. The HsdS subunit is required for DNA recognition specifying the target sequence, HsdM carries out the methyltransfer reaction, while HsdR carries out the DNA cleavage (Gormley *et al.*, 2001). In the present study, while *hsdR* (XCC0462) was induced in the L-vir strain, XCC2904, coding for a homolog of HsdS was repressed. XCC2732 and XCC2733, both related to the type IV RM system, were

induced in L-vir. XCC2732 codes for a conserved hypothetical protein described as 5-methylcytosine-specific restriction enzyme C, while XCC2733 codes for a McrB-like protein (5-methylcytosine-specific restriction enzyme B). The enzymatic complex McrBC, is a complex GTP(guanosine triphosphate)-requiring endonuclease, part of a type IV modification-dependent restriction system, that recognizes foreign DNA that contains methylated bases in a particular sequence (Stewart *et al.*, 2000; Loenen & Raleigh, 2014).

recJ (XCC1846), coding for a single-stranded DNA exonuclease, was also repressed in *Xcc* L-vir. *recJ* has been implicated in the regulation of *rpf* – *r*egulators of *p*athogenicity *f*actors – genes, which coordinate the regulation of genes involved in the biosynthesis of EPS and extracellular enzymes (Dow *et al.*, 2000).

XCC3363, coding for a conserved hypothetical protein with 4 transmembrane domains was also repressed. Despite its classification points it as a part of "defense mechanisms", no further information was available. Still, this protein appears to be conserved in *Xcc* strains, as well as in *Xylella fastidiosa*.

Differentially expressed genes involved in cell envelope and structure

Among the set of identified DEGs, nine encode for proteins involved in cell envelope and structure (Figure 29). Seven of them were repressed in *Xcc* L-vir strain, while the remaining two were induced in this strain.

The inner membrane protein coding genes *creD* (XCC2692) was found to be down-regulated in L-vir strain. *creD* is a conserved gene that encodes a predicted inner-membrane protein and is located near the *creBC* loci (Huang *et al.*, 2015), as previously mentioned. A *creD* deletion mutant of *Stenotrophomonas maltophilia* displayed a filamentous morphology – constriction appeared to begin at the septal ring but could not be completed – indicating a defect in constriction and separation. Morphology modifications in this mutant suggest that CreD is involved in the maintenance of cell envelope integrity. In *E. coli* and *Pseudomonas aeruginosa*, activation of CreBC increases *creD* expression; therefore, *creD* expression is often used as a measure of CreBC activation in these organisms. However, in *S. maltophilia* (a member of the Xanthomonadaceae family), CreBC TCS negatively regulates the promoter activity of *creD* (P_{creD}). Therefore, the repression of *creD* expression in *Xcc* L-vir strain may be a result of CreBC activation.

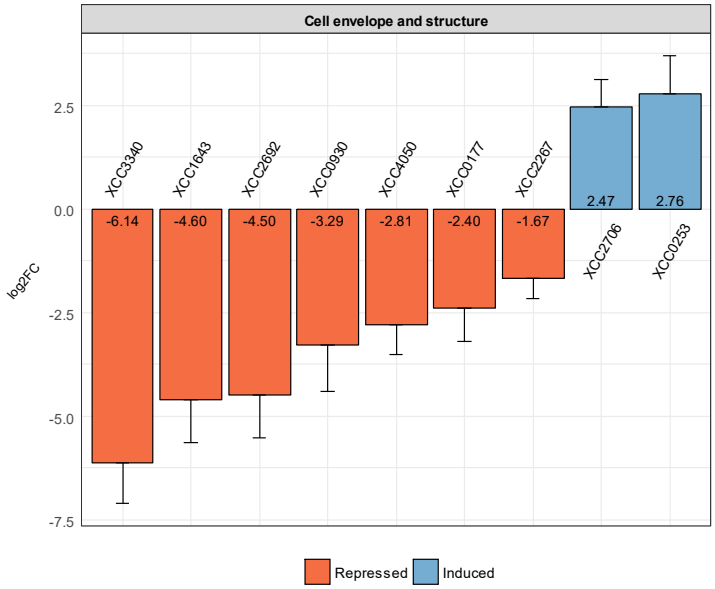


Figure 30 - Differentially expressed genes (DEGs) involved in cell envelope and structure. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

Additionally, XCC4050, coding for an ankyrin-like protein, was repressed in L-vir. Ankyrin (Ank) repeats, the most common protein-protein interaction motifs in eukaryotes, have been found in many microbes (Al-Khodor *et al.*, 2010). Although their presence in symbiotic bacterial and viral genomes was initially proposed as a result of horizontal gene transfer, evidence of these motifs in free-living organisms suggests a more ancient origin (Al-Khodor *et al.*, 2010). In *X. axonopodis* pv. *phaseoli*, the causative agent of common blight in bean, the protein coding gene *ankA* is located downstream of *katA*, coding for a catalase responsible for 90% of all catalase activity in this pathogen (Chauvatcharin *et al.*, 2005). Inactivation of *ankA* significantly reduced resistance to H_2O_2 in this pathogen and although the details of the functional role of AnkA in detoxification of peroxide are unclear, it was proposed that this Ank-repeat containing protein may be involved in the export of KatA (Chauvatcharin *et al.*, 2005). A similar catalase regulation mechanism has also been described for other bacterial pathogens, such as *P. aeruginosa* (Voronin & Kiseleva, 2007).

gtrB (XCC1643) and XCC3340, both coding for glycosyl transferase-like proteins, were highly repressed in *Xcc* L-vir, suggesting that these genes may play an important role in virulence related cellular processes. Glycosyl transferases (GTs) are ubiquitous enzymes that catalyze the synthesis of glycoside linkages by the transfer of a sugar residue from a donor substrate to an acceptor (Schmid *et al.*, 2016). More recently, GTs have been included in the 'bacterial toxin' category, since many of these enzymes are capable of glycosylating host proteins during pathogenesis, thus manipulating several hosts' cellular processes (Jank *et al.*, 2015). In *Xanthomonas* spp. GTs have been linked to LPS and EPS biosynthesis and biofilm formation, as well as adaptation to plant microenvironments, being required for full virulence (Patil *et al.*, 2007; Li *et al.*, 2012). Our results are in accordance with these findings,

since the least virulent strain expressed GTs to a lesser extent. The ortholog of XCC1643 in *Xcc* strain 8004 (XC_2588) was induced when the strain was grown in MMX medium (Liu *et al.*, 2013).

Two copies of *cls*, coding for cardiolipin synthase were found to have different expression patterns. XCC0177 was down-regulated in *Xcc* L-vir, while XCC2706 was up-regulated, evidencing the redundancy of genes involved in basic cellular structure. The glycerophospholipid cardiolipin is a membrane constituent characteristic of prokaryotes, playing an important role in bacterial stability and function, mostly through the formation and maintenance of dynamic protein–lipid and protein–protein interactions (Mileykovskaya & Dowhan, 2009; Serricchio & Bütikofer, 2012)

XCC2267, coding for a hypothetical protein containing SQS_PSY domains, found in Squalene/phytoene synthases, was also repressed. In bacteria, squalene serves as a precursor for pentacyclic hopanoids, such as hopane. Hopanoids are the bacterial substitutes of eukaryotic sterols, which have the capacity to modulate the ordering of lipids allowing cells to control membrane fluidity without compromising integrity (Sáenz *et al.*, 2012; van der Donk, 2015).

XCC0930, coding for a hypothetical protein that contains a domain like the one found on the cell wall hydrolase SleB, was also repressed in *Xcc* L-vir strain. SleB is found in *Bacillus subtilis* expressed during sporulation in an inactive form and deposited on the cell outer cortex (Moriyama *et al.*, 1996). When *Xcc* strain ATCC33913^T was grown in the presence of cabbage xylem sap, this gene was also repressed (de Bernonville *et al.*, 2014).

XCC0253, coding for a hypothetical protein containing carboxypeptidase regulatory like domains, was induced in *Xcc* L-vir strain. This protein is orthologous to a putative secreted protein in *Xanthomonas vesicatoria* (XCV0280), the causative agent of leaf spot of tomato and pepper.

DEGs involved in energy, carbon and central intermediary metabolism

Among the DEGs identified in this study, nine were involved in energy, carbon and central intermediary metabolism (Figure 30).

Of the eight DEGs involved in energy and carbon metabolism, *mocA* (XCC2671), coding for an oxidoreductase, *zwf* (XCC2136) coding for a glucose-6-phosphate 1-dehydrogenase, *aceE* (XCC3625) coding for pyruvate dehydrogenase subunit E1, and XCC0308 coding for a dehydrogenase, were up-regulated in L-vir. Contrarily, XCC3187, coding for a Indigoidine synthesis like protein, XCC1669 coding for a cytochrome like B561 protein, as well as XCC3695 and XCC2538, both coding for oxidoreductases, were found to be down-regulated in L-vir.

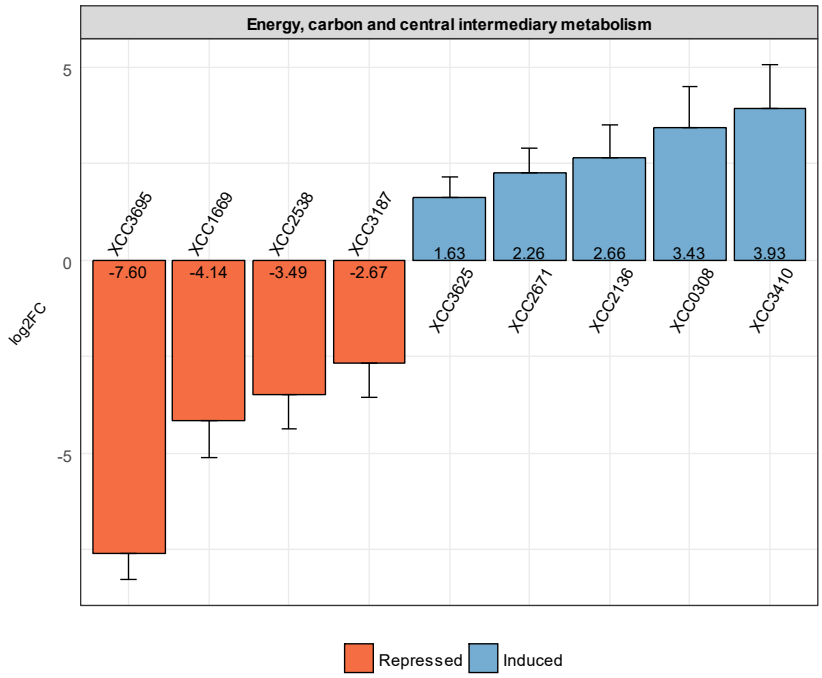


Figure 31 - Differentially expressed genes (DEGs) involved in energy, carbon and central intermediary metabolism. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

XCC3695 was the most down-regulated of the DEGs identified. Homologous proteins have also been described in *X. axonopodis* pv. *citri* strain 306 (da Silva *et al.*, 2002), *X. euvesicatoria* strain 85-10 (Thieme *et al.*, 2005) and in *Xylella fastidiosa* (Simpson *et al.*, 2000), suggesting that it may be a conserved protein among the Xanthomonodaceae family. XCC3695 was reported to be induced when *Xcc* strain 8004 was grown in the minimal medium MMX, along with several other ROS detoxification proteins, suggesting that it plays an important role in adaptation to harsh environments (Liu *et al.*, 2013).

nagA (XCC3410), coding for a N-acetylglucosamine-6-phosphate deacetylase, was up-regulated in L-vir. In *Xcc* strains, *nagA* belongs to a *c*arbohydrate *u*tilization loci containing *T*onB-dependent transporters (CUT loci) devoted to scavenging specific carbohydrates, playing an important role as part of the cytoplasmic pathway responsible for N-acetylglucosamine (GlcNAc) catabolism (Boulanger *et al.*, 2010). *Xcc nagA* mutants were shown to display growth impairment in the presence of N-acetylglucosamine and a reduction of 90-95% in pathogenicity against *A. thaliana* when compared to wild-type strains (Boulanger *et al.*, 2014).

DEGs involved in protein degradation and post-translational modifications

A total of eight genes involved in protein degradation and post-translational modifications were identified, corresponding to 6% of the global DEG set (Figure 31).

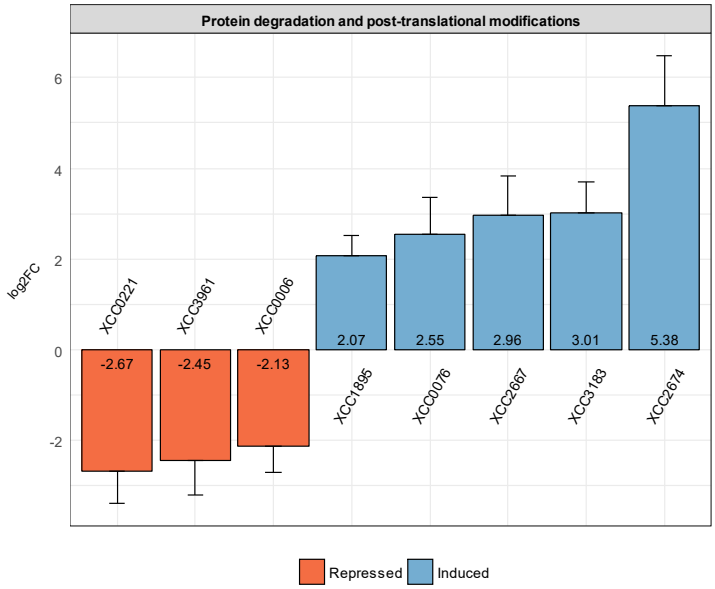


Figure 32 - Differentially expressed genes (DEGs) involved in protein degradation and post-translational modifications. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

In several animal/human bacterial pathogens, the involvement of secreted proteases in pathogenicity-related processes includes the maturation of virulence factors, the processing of host defensive proteins during infection and in biofilm formation (Molière & Turgay, 2013).

Proteases also seem to play a central role in plant-pathogen interactions, since many of the described bacterial effectors and plant resistance proteins have been found to be proteases (Xia, 2004). Post-translational modification of host proteins through proteolytic processing is a widely used mechanism in regulating the plant defense response, present in several plant pathogens (Xia, 2004).

Of the DEGs involved in protein degradation, XCC2674 and XCC2667, coding for two extracellular serine proteases, and XCC0076, coding for a zinc metalloprotease, were induced. On the other hand, XCC3961, coding for a dipeptidyl peptidase IV, and XCC0221, coding for an acetyltransferase, were repressed in L-vir strain.

It is important to notice that XCC2667 and XCC0076 were both reported to be induced when *Xcc* strain 8004 was grown in MMX, while the latter was also induced in the presence of xylem sap (for *Xcc* strain ATCC33913^T), suggesting that it may have a significant role during pathogenesis (Liu *et al.*, 2013; de Bernonville *et al.*, 2014). On the other hand, XCC0221 was reported to be induced in *Xcc* strain 8004 during growth on MMX (Liu *et al.*, 2013). The differential expression of extracellular proteases between L-vir and H-vir strains are therefore in line with the key role play in quorum sensing mechanisms necessary for environmental adaptation (Waters & Bassler, 2005).

Additionally, three genes coding for poorly characterized proteins putatively involved in protein modifications, were identified as differentially expressed. XCC0006, coding for a conserved

hypothetical protein with homology to a putative Zn-dependent protease, was repressed in L-vir. On the other hand, XCC3183, coding for a hypothetical protein containing Phospholipid methyltransferase and isoprenylcysteine carboxyl methyltransferase (ICMT) family domains, and XCC1895, coding for an uncharacterized protein were induced in *Xcc* L-vir strain.

DEGs involved in the biosynthesis of amino acids, cofactors, prosthetic groups and carriers

A total of six genes encoding for proteins involved in the biosynthesis of amino acids, cofactors, prosthetic groups and carriers were found to be differentially expressed (Figure 32). Four of them were repressed in the least virulent strain (L-vir), while only one was induced.

ilvE (XCC0850) coding for a branched-chain amino acid (BCAA) aminotransferase, and *ilvG* (XCC3324) coding for acetolactate synthase 2 catalytic large subunit, were both down-regulated in *Xcc* L-vir strain.

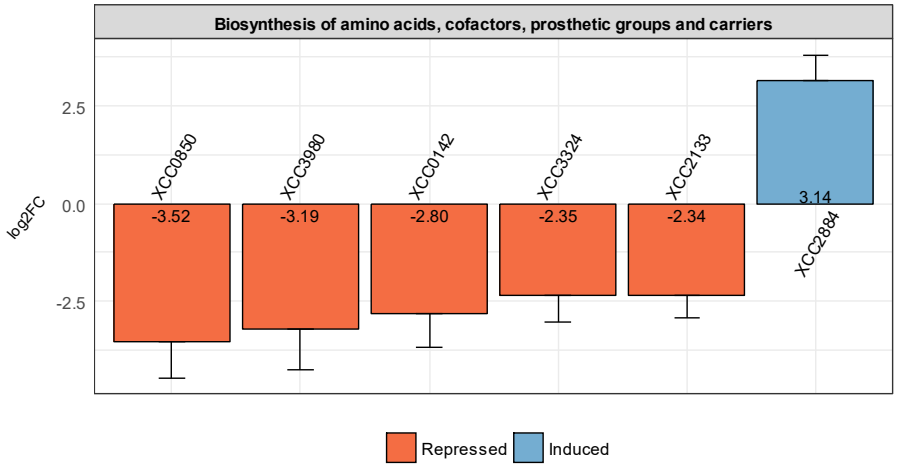


Figure 33 - Differentially expressed genes (DEGs) involved in the biosynthesis of amino acids, cofactors, prosthetic groups and carriers. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

ilvGMEDA operon is involved in the biosynthesis of the BCAAs valine, leucine and isoleucine (Vitreschak *et al.*, 2004). In plants, the BCAAs allosterically regulate the activity of acetolactate synthase (ALS). Accumulation of BCAA isoleucine is correlated with the development of "Frenching disease" in tobacco plants (Steinberg *et al.*, 1950). In "Frenching disease," overproduction of isoleucine by the saprophytic bacteria inhibits the activity of ALS in the tobacco, shutting down synthesis of valine and leucine, which in turn disrupted essential protein metabolism, leading to chlorosis and cell death (Steinberg *et al.*, 1950). By using pathogens excreting large amounts of BCAAs, a strategy was created to control weed development (Thompson *et al.*, 2007). Although a direct effect of BCAAs on pathogen resistance has not yet been reported, a recent study indicates that fluctuations of isoleucine and valine metabolism can affect plant defense (Zeier, 2013). The repression of essential BCAA biosynthesis genes

in *Xcc* L-vir, likely leading to diminished BCAA accumulation in plant tissue, could ultimately result in decreased virulence, due to its inability to inhibit the activity of ALS.*ilvG* has been reported to be induced when *Xcc* strain 8004 was grown in the T3SS activating medium MMX, suggesting that activation of amino acid biosynthetic pathways occur in response to the nutrient starvation this medium imposes, when compared to the growth in a rich medium (Liu *et al.*, 2013). Although our results cannot be directly compared to these, the modulation of branched-chain amino acid biosynthetic pathways during pathogenesis appears to be clear.

XCC2133 (*ygfZ*), encoding for a putative aminomethyl transferase, responsible for tRNA modification, was found repressed in *Xcc* L-vir strain. Similarly, this locus was found to be repressed when *Xcc* strain ATCC33913^T was grown in the presence of cabbage xylem sap (de Bernonville *et al.*, 2014). XCC0142, a conserved hypothetical protein containing lipase-like domains (Lipase_GDSL domains), was also repressed in L-vir strain.

apbA (XCC3980), coding for a 2-dehydropantoate 2-reductase, involved in pantothenate and CoA biosynthesis from precursors valine/L-aspartate, was found to be down-regulated in L-vir.

bioA (XCC2884), coding for adenosylmethionine-8-amino-7-oxononanoate aminotransferase, was found to be up-regulated in L-vir. The *bio* operon is involved in the biosynthesis of biotin, an important cofactor incorporated in several biotin-dependent carboxylases, present in prokaryotes and eukaryotes (Streit & Entcheva, 2003).

DEGs involved in transcription and translation

Of the total DEG set, four genes are involved in transcription or translation-related processes (Figure 33).

The anticodon nuclease coding gene XCC0464, involved in RNA degradation, was found to be induced in L-vir strain. XCC1463, coding for a hypothetical protein containing a 'Prophage CP4-57 regulatory protein' domain, was found repressed. In *E. coli* strain K12, deletion of prophage CP4-57 has been associated to increased biofilm formation, as well as over-expression of *flg*, *flh* and *fli* motility operons and repression of key enzymes involved in both the Tri-Carboxylic Acid cycle and lactate utilization (Wang *et al.*, 2009).

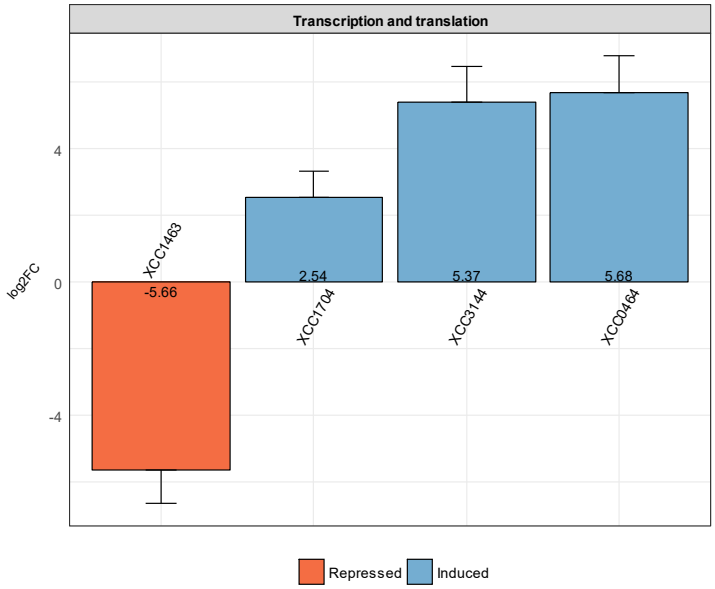


Figure 34 - Differentially expressed genes (DEGs) involved in transcription and translation. DEG expression was classified as repressed or induced, according to its log_2FC (indicated for each bar) calculated on the basis of L-vir/H-vir expression. XCC(…) – corresponding locus, based on the annotation for *Xanthomonas campestris* pv. *campestris* ATCC33913^T.

XCC3144, coding for a hypothetical protein containing a LysR substrate binding domain was found induced in L-vir strain. The LysR-type of transcriptional regulators are the most common transcriptional regulators in bacteria, involved in modulation of several genes in metabolism, cell division, quorum sensing, virulence, motility, nitrogen fixation, oxidative stress responses, toxin production, attachment and secretion, among other cellular processes (Maddocks & Oyston, 2008).

XCC1704, coding for TruD protein, a tRNA pseudouridine synthase, was also found to be up-regulated in *Xcc* L-vir. TruD is responsible for the isomerization of uridine to pseudouridine, the most common modification seen in transfer, ribosomal and splicing RNAs (Kaya *et al.*, 2004).

Poorly characterized differentially expressed gene clusters

Among the 154 DEGs identified, 41 of them correspond to poorly characterized genes, representing 26% of the global DEG set (Figure 23).

Of these genes, XCC0030 and XCC0113 were previously reported by other authors as involved in pathogenicity. XCC0030, highly repressed in *Xcc* L-vir strain had also been identified as induced in MMX, the minimal medium thought to induce expression of T3SS (Liu *et al.*, 2013). XCC0113, also induced in *Xcc* L-vir strain, was slightly repressed when *Xcc* strain ATCC33913^T cells were grown in the presence of cabbage xylem sap (de Bernonville *et al.*, 2014).

Among poorly characterized genes that were differentially expressed between both *Xcc* strains used in this study, four clusters of closely located genes were identified.

Clusters XCC0536-XCC0538, XCC1133-XCC1134 and XCC3934-3936 were repressed in *Xcc* L-vir, while cluster XCC2091-XCC2096 (including the DNA helicase coding gene XCC2095) was induced. Cluster XCC0536-XCC0538 is near XCC0530, the coding gene for XopX, a candidate type III effector, also repressed. Downstream locus XCC0539 codes for the outer membrane protein OmpW, known to be involved in the transport of small hydrophobic compounds in *E. coli* and in the degradation of alkanes in *Pseudomonas oleovorans* (Hong *et al.*, 2006). Cluster XCC3934-XCC3936 is located near XCC3932, coding for an AraC-family transcriptional regulator, which was also found repressed in L-vir. Interestingly, the downstream genes XCC3937, coding for a pantothenic acid kinase, and XCC3938, coding for a bifunctional transcriptional repressor of the biotin operon/biotin acetyl-CoA-carboxylase synthase, are also described as transcriptional regulators. Except for XCC2092, coding for an oxygen-independent coproporphyrinogen III oxidase, cluster XCC2091-XCC2096 was induced in *Xcc* L-vir. This cluster includes the DNA helicase-like protein coding locus XCC2095, and XCC2094 coding for a tannase precursor involved in degradation of plant tannins.

Concluding remarks

While most studies have focused on *in vitro* cultures of well-known *Xcc* strains in conditions mimicking infection, the work here presented corresponds to a novel tactic on understanding the transcriptome of the *Xcc* – *B. oleracea* pathosystem. Using an *in vivo* approach, the actual transcriptional reprogramming that convoys the infection was accurately disclosed and the results obtained allowed to clarify some of the questions concerning pathogenicity/virulence of *Xcc*.

From the start, this work made clear that for equally pathogenic strains, sharing pathovar and race, virulence can be contrasting. It was also evident that these differences are host-independent, highlighting the importance of features intrinsic to the pathogen for disease onset and progression and thus fading the presumed effects of host varieties and cultivars.

With those premises in mind, the transcriptome profiling carried out in this study allowed to identify a substantial set of differentially expressed genes between two *Xcc* strains showing contrasting virulence. As expected, a large proportion of the identified genes were involved in pathogenicity and adaptation. Within this subset, a similar number of genes was either induced or repressed in both strains, suggesting that there is no global repression or induction patterns that could account for the low or high virulence of the strain. However, a total of five Type III effectors were among this set – XopE2 and XopD were associated with a low-virulence phenotype, while XopAC, XopX and XopR were associated with a high-virulence phenotype. It is also important to mention, that except for XopE2, all other Type III effectors were expressed in both strains, indicating that the simple presence of avirulence/virulence proteins is not sufficient to explain contrasting virulence phenotypes.

Several other genes, known to be involved in pathogenicity and adaptation functions were also identified, and found to be repressed in the lowest virulent strain. Surprisingly, XCC3695 coding for a poorly characterized oxidoreductase, was highly induced in the highest virulent strain. These findings suggest that reactive oxygen species detoxification may play a very important role in virulence regulation, and that this gene should be further evaluated from a functional stand-point.

Although none of the known major regulators of *Xanthomonas* pathogenicity, such as HrpX, were identified as differentially expressed, other transcriptional regulators were repressed in L-vir, as well as some important sensory mechanisms. Combined, these findings suggest that while the expression of major regulators are involved in pathogenicity, an additional set of transcriptional regulators may be involved in the fine-tuning of virulence.

Globally, the results obtained in this study suggest that the low virulence phenotype is the combined result of impaired sensory mechanisms, reduced detoxification of ROS, decreased motility, higher production of PAMPs and an overexpression of avirulence proteins.

Contrastingly, the highly virulent strain appears to be better equipped to evade the initial plant defense responses, whether by staying under the host 'radar' avoiding detection, or by the ability to counteract those responses when detected. On the other hand, upon detection, highly virulent pathogens will show a decreased expression of avirulence proteins, making them less recognizable by the hosts ETI mechanisms, and thus able to continue multiplying and cause more severe disease symptoms.

Highly virulent pathogens correspond to the extreme of the virulence continuum, causing severe losses due to host fitness reduction, and have thus been the focus of plant pathologists and breeders. However, the identification of avirulence genes and matching resistance genes is a difficult and time-consuming process and, for *Xcc-B. oleracea* pathosystem, it has certainly been tenuous. Moreover, resistance based on pairs of avirulence-resistance genes is known to be rapidly overcome in modern agricultural crops, due to high pathogen diversity coupled with a limited host genetic background. An alternative strategy could involve the use of genetic tolerance, to minimize the impact of pathogens in cropping systems. Tolerance can be defined as the ability of a susceptible plant to endure pathogen damage without suffering a loss in yield (Barrett *et al.*, 2009). In order to understand the polygenic nature of tolerance, the study of less virulent pathogens is of utmost importance, unveiling what makes them tolerable and distinguishes them of their intolerable highly virulent counterparts.

References

Ailloud F, Lowe TM, Robène I, Cruveiller S, Allen C, Prior P, 2016. *In planta* comparative transcriptomics of host-adapted strains of *Ralstonia solanacearum*. *PeerJ* **4**, e1549.

Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y, 2010. Functional diversity of ankyrin repeats in microbial proteins. *Trends in microbiology* **18**, 132–139.

Anders S, Pyl PT, Huber W, 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169.

Anderson GG, Yahr TL, Lovewell RR, O'Toole GA, 2010. The *Pseudomonas aeruginosa* magnesium transporter MgtE inhibits transcription of the type III secretion system. *Infection and immunity* **78**, 1239–1249.

Arnold DL, Jackson RW, 2011. Bacterial genomes: evolution of pathogenicity. *Current opinion in plant biology* **14**, 385–391.

Arnold DL, Pitman A, Jackson RW, 2003. Pathogenicity and other genomic islands in plant pathogenic bacteria. *Molecular Plant Pathology* **4**, 407–420.

Babraham Bioinformatics FastQC A Quality Control tool for High Throughput Sequence Data.

Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J, 2009. Continua of specificity and virulence in plant host-pathogen interactions: Causes and consequences. *New Phytologist* **183**, 513–529.

Bent AF, Mackey D, 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology* **45**, 399–436.

de Bernonville TD, Noël LD, Sancristobal M *et al.*, 2014. Transcriptional reprogramming and phenotypical changes associated with growth of *Xanthomonas campestris* pv. *campestris* in cabbage xylem sap. *FEMS microbiology ecology* **89**, 527–541.

Blanvillain S, Meyer D, Boulanger A *et al.*, 2007. Plant Carbohydrate Scavenging through TonB-Dependent Receptors: A Feature Shared by Phytopathogenic and Aquatic Bacteria. *PloS one* **2**, e224.

Boulanger A, Déjean G, Lautier M *et al.*, 2010. Identification and regulation of the N-acetylglucosamine utilization pathway of the plant pathogenic bacterium *Xanthomonas campestris* pv. *campestris*. *Journal of bacteriology* **192**, 1487–97.

Boulanger A, Zischek C, Lautier M *et al.*, 2014. The plant pathogen *Xanthomonas campestris* pv.

campestris exploits N-acetylglucosamine during infection. *mBio* **5**, e01527-14.

Büttner D, Bonas U, 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* **34**, 107–133.

Camilios-Neto D, Bonato P, Wassem R *et al.*, 2014. Dual RNA-seq transcriptional analysis of wheat roots colonized by Azospirillum brasilense reveals up-regulation of nutrient acquisition and cell cycle genes. *BMC genomics* **15**, 1–13.

Chao N-X, Wei K, Chen Q *et al.*, 2008. The rsmA-like gene rsmA(Xcc) of *Xanthomonas campestris* pv. *campestris* is involved in the control of various cellular processes, including pathogenesis. *Molecular plant-microbe interactions : MPMI* **21**, 411–423.

Chauvatcharin N, Atichartpongkul S, Utamapongchai S, Whangsuk W, Vattanaviboon P, Mongkolsuk S, 2005. Genetic and physiological analysis of the major OxyR-regulated katA from *Xanthomonas campestris* pv. *phaseoli*. *Microbiology* **151**, 597–605.

Choi J, Shin D, Yoon H *et al.*, 2010. *Salmonella* pathogenicity island 2 expression negatively controlled by EIIANtr-SsrB interaction is required for *Salmonella* virulence. *Proceedings of the National Academy of Sciences* **107**, 20506–20511.

Cicero DO, Contessa GM, Pertinhez TA *et al.*, 2007. Solution structure of ApaG from *Xanthomonas axonopodis* pv. *citri* reveals a fibronectin-3 fold. *Proteins* **67**, 490–500.

Coburn PS, Gilmore MS, 2003. The *Enterococcus faecalis* cytolysin: A novel toxin active against eukaryotic and prokaryotic cells. *Cellular Microbiology* **5**, 661–669.

Cubero J, Graham JH, 2004. The leucine-responsive regulatory protein (*lrp*) gene for characterization of the relationship among *Xanthomonas* species. *International Journal of Systematic and Evolutionary Microbiology* **54**, 429–437.

Déjean G, Blanvillain-Baufumé S, Boulanger A *et al.*, 2013. The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv. *campestris* controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytologist* **198**, 899–915.

Deng W, Li C, Xie J, 2013. The underling mechanism of bacterial TetR/AcrR family transcriptional repressors. *Cellular Signalling* **25**, 1608–1613.

van der Donk WA, 2015. Bacteria Do It Differently: An Alternative Path to Squalene. *ACS central science* **1**, 64–65.

Dow JM, Daniels MJ, Feng J-X, Tang J-L, Barber CE, 2000. Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*.

Microbiology **146**, 885–891.

Dunger G, Llontop E, Guzzo CR, Farah CS, 2016. The Xanthomonas type IV pilus. *Current Opinion in Microbiology* **30**, 88–97.

F.Wendisch V, 2007. *Amino Acid Biosynthesis – Pathways, Regulation andMetabolic Engineering* (A Steinbüchel, Ed,). Münster: Springer - Verlag.

Fargier E, Manceau C, 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**, 805–818.

Flor HH, 1956. The Complementary Genic Systems in Flax and Flax Rust. *Advances in Genetics* **8**, 29–54.

Fuangthong M, Jittawuttipoka T, Wisitkamol R *et al.*, 2015. IscR plays a role in oxidative stress resistance and pathogenicity of a plant pathogen, *Xanthomonas campestris*. *Microbiological Research* **170**, 139–146.

Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL, 1997. Arac/XylS family of transcriptional regulators. *Microbiology and molecular biology reviews : MMBR* **61**, 393–410.

Galperin MY, 2004. Bacterial signal transduction network in a genomic perspective. *Environmental Microbiology* **6**, 552–567.

Galperin MY, 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC microbiology* **5**.

Geiss-friedlander R, Melchior F, 2007. Concepts in sumoylation: a decade on. *Nature Reviews Molecular Cell Biology* **8**, 947–956.

Gibson MM, Bagga DA, Miller CG, Maguire ME, 1991. Magnesium transport in Salmonella typhimurium: the influence of new mutations conferring Co2+ resistance on the CorA Mg2+ transport system. *Molecular microbiology* **5**, 2753–2762.

Goel AK, Rajagopal L, Nagesh N, Sonti R V, 2002. Genetic locus encoding functions involved in biosynthesis and outer membrane localization of xanthomonadin in *Xanthomonas oryzae* pv. *oryzae*. *Journal of bacteriology* **184**, 3539–3548.

Gormley N a, Gormley N a, Watson M a, Watson M a, Halford SE, Halford SE, 2001. Bacterial Restriction– Modification Systems. In: *eLS*.

Gouin A, Legeai F, Nouhaud P, Whibley A, Simon J-C, Lemaitre C, 2015. Whole-genome re-sequencing

of non-model organisms: lessons from unmapped reads. *Heredity* **114**, 494–501.

Gurian-Sherman D, Lindow SE, 1993. Bacterial ice nucleation: significance and molecular basis. *FASEB journal* **7**, 1338–1343.

Guy E, Lautier M, Chabannes M *et al.*, 2013. xopAC-triggered Immunity against *Xanthomonas* Depends on *Arabidopsis* Receptor-Like Cytoplasmic Kinase Genes PBL2 and RIPK (C-H Yang, Ed,). *PLoS ONE* **8**, e73469.

He Y-W, Ng AY-J, Xu M *et al.*, 2007a. *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Molecular Microbiology* **64**, 281–292.

He Y-Q, Zhang L, Jiang B-L *et al.*, 2007b. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome biology* **8**, R218.

Hong H, Patel DR, Tamm LK, van den Berg B, 2006. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *The Journal of biological chemistry* **281**, 7568–7577.

Huang HH, Lin YT, Chen WC, Huang YW, Chen SJ, Yang TC, 2015. Expression and functions of CreD, an inner membrane protein in *Stenotrophomonas maltophilia*. *PLoS ONE* **10**, 1–16.

Huang DL, Tang DJ, Liao Q *et al.*, 2008. The Zur of *Xanthomonas campestris* functions as a repressor and an activator of putative zinc homeostasis genes via recognizing two distinct sequences within its target promoters. *Nucleic Acids Research* **36**, 4295–4309.

Jacob TR, de Laia ML, Moreira LM *et al.*, 2014. Type IV Secretion System Is Not Involved in Infection Process in Citrus. *International journal of microbiology* **2014**, 763575.

Jacques M-A, Arlat M, Boulanger A *et al.*, 2016. Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*: French Network on Xanthomonads (FNX). *Annual Review of Phytopathology* **54**, 163–187.

Jank T, Belyi Y, Aktories K, 2015. Bacterial glycosyltransferase toxins. *Cellular Microbiology* **17**, 1752–1765.

Jones JDG, Dangl JL, 2006. The plant immune system. *Nature* **444**, 323–329.

Kaya Y, Del Campo M, Ofengand J, Malhotra A, 2004. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *The Journal of biological chemistry* **279**, 18107–18110.

Khan A, Mathelier A, 2017. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287.

Kim J-G, Stork W, Mudgett MB *et al.*, 2013. *Xanthomonas* type III effector XopD desumoylates tomato transcription factor SlERF4 to suppress ethylene responses and promote pathogen growth. *Cell host & microbe* **13**, 143–154.

Kim J-GJ-G, Taylor KW, Hotson A, Keegan M, Schmelz EA, Mudgett MB, 2008. XopD SUMO Protease Affects Host Transcription, Promotes Pathogen Growth, and Delays Symptom Development in <i>Xanthomonas<-Infected Tomato Leaves. *The Plant Cell* **20**, 1915–1929.

Kimbrel J a., Di Y, Cumbie JS, Chang JH, 2011. RNA-Seq for Plant Pathogenic Bacteria. *Genes* **2**, 689–705.

Koebnik R, Krüger A, Thieme F, Urban A, Bonas U, 2006. Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *Journal of bacteriology* **188**, 7652–7660.

Kuo MMC, Saimi Y, Kung C, 2003. Gain-of-function mutations indicate that *Escherichia coli* Kch forms a functional K+ conduit in vivo. *EMBO Journal* **22**, 4049–4058.

Kvint K, Nachin L, Diez A, Nyström T, 2003. The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology* **6**, 140–145.

Lemaire C, Hajri A, Brin C, Hunault G, 2009. A « Repertoire for Repertoire » Hypothesis : Repertoires of Type Three Effectors are Candidate Determinants of Host Specificity in *Xanthomonas*. *PloS one* **4**, e6632.

Li H, Durbin R, 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595.

Li C, Tao J, Mao D *et al.*, 2011. A Novel Manganese Efflux System, YebN, Is Required for Virulence by *Xanthomonas oryzae* pv. *oryzae* (MR Volkert, Ed,). *PLoS ONE* **6**, e21983.

Li J, Wang N, Vauterin L *et al.*, 2012. The *gpsX* gene encoding a glycosyltransferase is important for polysaccharide production and required for full virulence in *Xanthomonas citri* subsp. *citri*. *BMC Microbiology* **12**, 31.

Liu X, Grabherr HM, Willmann R *et al.*, 2014. Host-induced bacterial cell wall decomposition mediates pattern-triggered immunity in *Arabidopsis*. *eLife* **3**, 1544–1556.

Liu W, Yu Y-HH, Cao S-YY *et al.*, 2013. Transcriptome profiling of *Xanthomonas campestris* pv. *campestris* grown in minimal medium MMX and rich medium NYG. *Research in Microbiology* **164**,

466–479.

Loenen WAM, Raleigh EA, 2014. The other face of restriction : modification-dependent enzymes. **42**, 56–69.

Love MI, Huber W, Anders S *et al.*, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.

Lu H, Patil P, Van Sluys M-A *et al.*, 2008. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PloS one* **3**, e3828.

Lu J, Zheng Y, Yamagishi H *et al.*, 2003. Motif CXCC in nitrile hydratase activator is critical for NHase biogenesis in vivo. *FEBS Letters* **553**, 391–396.

Maddocks SE, Oyston PCF, 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**, 3609–3623.

Meibom KL, Charbit A, 2010. *Francisella tularensis* metabolism and its relation to virulence. *Frontiers in microbiology* **1**, 1–13.

Metz M, Dahlbeck D, Morales CQ, Al Sady B, Clark ET, Staskawicz BJ, 2005. The conserved Xanthomonas campestris pv. vesicatoria effector protein XopX is a virulence factor and suppresses host defense in Nicotiana benthamiana. *Plant Journal* **41**, 801–814.

Mileykovskaya E, Dowhan W, 2009. Cardiolipin membrane domains in prokaryotes and eukaryotes. *Biochimica et biophysica acta* **1788**, 2084–91.

Mole BM, Baltrus D a, Dangl JL, Grant SR, 2007. Global virulence regulation networks in phytopathogenic bacteria. *Trends in microbiology* **15**, 363–71.

Molière N, Turgay K, 2013. *Regulated Proteolysis in Microorganisms*.

Moriyama R, Hattori A, Miyata S, Kudoh S, Makino S, 1996. A gene (sleB) encoding a spore cortex-lytic enzyme from Bacillus subtilis and response of the enzyme to L-alanine-mediated germination. *Journal of bacteriology* **178**, 6059–6063.

Nimchuk ZL, Fisher EJ, Desveaux D, Chang JH, Dangl JL, 2007. The HopX (AvrPphE) family of *Pseudomonas syringae* type III effectors require a catalytic triad and a novel N-terminal domain for function. *Mol Plant Microbe Interact* **20**, 346–357.

Noinaj N, Guillier M, Barnard TJ, Buchanan SK, 2010. TonB-dependent transporters: regulation, structure, and function. *Annual review of microbiology* **64**, 43–60.

Nyström T, Neidhardt FC, 1994. Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol Microbiol* **11**, 537–544.

Pastrik K-H, 2000. Detection of *Clavibacter michiganensis* subsp. *sepedonicus* in Potato Tubers by Multiplex PCR with Coamplification of Host DNA. *European Journal of Plant Pathology* **106**, 155–165.

Pastrik K-H, Elphinstone JG, Pukall R, 2002. Sequence Analysis and Detection of *Ralstonia solanacearum* by Multiplex PCR Amplification of 16S–23S Ribosomal Intergenic Spacer Region with Internal Positive Control. *European Journal of Plant Pathology* **108**, 831–842.

Patil PB, Bogdanove AJ, Sonti R V *et al.*, 2007. The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in xanthomonads that infect rice, citrus and crucifers. *BMC Evolutionary Biology* **7**, 243.

Pfeilmeier S, Caly DL, Malone JG, 2016. Bacterial pathogenesis of plants : future challenges from a microbial perspective. *Molecular plant pathology* **17**, 1298–1313.

Popov G, Fraiture M, Brunner F, Sessa G, 2016. Multiple *Xanthomonas euvesicatoria* Type III Effectors Inhibit flg22-Triggered Immunity. *Molecular plant-microbe interactions : MPMI* **29**, 651–660.

Qian W, Han Z-J, He C, 2008. Two-component signal transduction systems of *Xanthomonas* spp.: a lesson from genomics. *Molecular plant-microbe interactions : MPMI* **21**, 151–161.

Qian W, Jia Y, Ren S-X *et al.*, 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome research* **15**, 757–767.

Rausher MD, 2001. Co-evolution and plant resistance to natural enemies. *Nature* **411**, 857–864.

Rigano LA, Payette C, Yun M *et al.*, 2007. Bacterial Cyclic b- (1 , 2) -Glucan Acts in Systemic Suppression of Plant Immune Responses. *The Plant Cell* **19**, 2077–2089.

Römling U, Galperin MY, Gomelsky M, 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiology and molecular biology reviews : MMBR* **77**, 1–52.

Roux B, Bolot S, Guy E *et al.*, 2015. Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. *BMC genomics* **16**, 1–19.

Ryan RP, Fouhy Y, Lucey JF *et al.*, 2007. Cyclic di-GMP signalling in the virulence and environmental adaptation of *Xanthomonas campestris*. *Molecular Microbiology* **63**, 429–442.

Sáenz JP, Sezgin E, Schwille P, Simons K, 2012. Functional convergence of hopanoids and sterols in membrane ordering. *Proceedings of the National Academy of Sciences* **109**, 14236–14240.

Sato M, Tsuda K, Wang L *et al.*, 2010. Network Modeling Reveals Prevalent Negative Regulatory Relationships between Signaling Sectors in *Arabidopsis* Immune Signaling (S He, Ed,). *PLoS Pathogens* **6**, e1001011.

Schauer K, Rodionov DA, de Reuse H *et al.*, 2008. New substrates for TonB-dependent transport: do we only see the "tip of the iceberg"? *Trends in biochemical sciences* **33**, 330–338.

Schmid J, Heider D, Wendel NJ, Sperl N, Sieber V, 2016. Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Frontiers in microbiology* **7**.

Serricchio M, Bütikofer P, 2012. An essential bacterial-type cardiolipin synthase mediates cardiolipin formation in a eukaryote. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E954-61.

da Silva ACR, Ferro JA, Reinach FC *et al.*, 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**, 459–463.

Simpson AJG, Reinach FC, Arruda P *et al.*, 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151–157.

Sinha D, Gupta MK, Patel HK, Ranjan A, Sonti R V., 2013. Cell Wall Degrading Enzyme Induced Rice Innate Immune Responses Are Suppressed by the Type 3 Secretion System Effectors XopN, XopQ, XopX and XopZ of *Xanthomonas oryzae* pv. *oryzae* (E Chaves-Olarte, Ed,). *PLoS ONE* **8**, e75867.

Solcan N, Kwok J, Fowler PW *et al.*, 2012. Alternating access mechanism in the POT family of oligopeptide transporters. *The EMBO Journal* **31**, 3411–3421.

Sondermann H, Shikuma NJ, Yildiz FH, 2012. You've come a long way: C-di-GMP signaling. *Current Opinion in Microbiology* **15**, 140–146.

Sousa V, Hey J, 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics* **14**, 404–414.

Souza DP, Oka GU, Alvarez-Martinez CE *et al.*, 2015. Bacterial killing via a type IV secretion system. *Nature communications* **6**, 6453.

Steinberg RA, Bowling JD, McMurtrey JE, Jr., 1950. Accumulation of free amino acids as a chemical basis for morphological symptoms in tobacco manifesting frenching and mineral deficiency symptoms. *Plant physiology* **25**, 279–288.

Stewart FJ, Panne D, Bickle TA, Raleigh EA, 2000. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *Journal of Molecular Biology* **298**, 611–622.

Stock AM, Robinson VL, Goudreau PN, 2000. Two-Component Signal Transduction. *Reactions* **69**, 183–215.

Stork W, Kim J-G, Mudgett MB, 2015. Functional Analysis of Plant Defense Suppression and Activation by the <i>Xanthomonas Core Type III Effector XopX. *Molecular plant-microbe interactions : MPMI* **28**, 180–194.

Streit WR, Entcheva P, 2003. Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Appl Microbiol Biotechnol* **61**, 21–31.

Tan CM, Li M-Y, Yang P-Y *et al.*, 2015. *Arabidopsis* HFR1 Is a Potential Nuclear Substrate Regulated by the *Xanthomonas* Type III Effector XopDXcc8004 (Y-W He, Ed,). *PLOS ONE* **10**, e0117067.

Tan H, West JA, Ramsay JP *et al.*, 2014. Comprehensive overexpression analysis of cyclic-di-GMP signalling proteins in the phytopathogen Pectobacterium atrosepticum reveals diverse effects on motility and virulence phenotypes. *Microbiology* **160**, 1427–1439.

Thaw P, Sedelnikova SE, Muranova T *et al.*, 2006. Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Research* **34**, 1439–1449.

Thieme F, Koebnik R, Bekel T *et al.*, 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium Xanthomonas campestris pv. vesicatoria revealed by the complete genome sequence. *Journal of bacteriology* **187**, 7254–7266.

Thieme F, Szczesny R, Urban A, Kirchner O, Hause G, Bonas U, 2007. New type III effectors from *Xanthomonas campestris* pv. *vesicatoria* trigger plant reactions dependent on a conserved N-myristoylation motif. *Molecular plant-microbe interactions : MPMI* **20**, 1250–1261.

Thompson BM, Kirkpatrick MM, Sands DC, Pilgeram AL, 2007. Pathogens for Control of Weeds. In: *Novel Biotechnologies for Biocontrol Agent Enhancement and Management*. Dordrecht: Springer Netherlands, 267–275.

Vicente JG, Taylor JD, Sharpe a G, Parkin I a P, Lydiate DJ, King GJ, 2002. Inheritance of Race-Specific Resistance to *Xanthomonas campestris* pv. *campestris* in Brassica Genomes. *Phytopathology* **92**, 1134–1341.

Vitreschak AG, Lyubetskaya E V., Shirshin MA, Gelfand MS, Lyubetsky VA, 2004. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: Comparative genomics analysis. *FEMS Microbiology Letters* **234**, 357–370.

Vorhölter FJ, Niehaus K, Pühler A, 2001. Lipopolysaccharide biosynthesis in *Xanthomonas campestris* pv. *campestris*: A cluster of 15 genes is involved in the biosynthesis of the LPS O-antigen and the

LPS core. *Molecular Genetics and Genomics* **266**, 79–95.

Voronin DA, Kiseleva E V., 2007. Functional role of proteins containing ankyrin repeats. *Cell and Tissue Biology* **49**, 989–999.

Vorwerk H, Mohr J, Huber C *et al.*, 2014. Utilization of host-derived cysteine-containing peptides overcomes the restricted sulphur metabolism of *Campylobacter jejuni*. *Molecular Microbiology* **93**, 1224–1245.

Wadhams GH, Armitage JP, 2004. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology* **5**, 1024–1037.

Wang X, Kim Y, Wood TK, 2009. Control and benefits of CP4-57 prophage excision in Escherichia coli biofilms. *The International Society for Microbial Ecology Journal* **3**, 1164–1179.

Wang S, Sun J, Fan F, Tan Z, Zou Y, Lu D, 2016. A *Xanthomonas oryzae* pv. *oryzae* effector, XopR, associates with receptor-like cytoplasmic kinases and suppresses PAMP-triggered stomatal closure. *Science China Life Sciences* **59**, 897–905.

Wang J-Y, Zhou L, Chen B *et al.*, 2015. A functional 4-hydroxybenzoate degradation pathway in the phytopathogen *Xanthomonas campestris* is required for full pathogenicity. *Scientific Reports* **5**, 18456.

Waters CM, Bassler BL, 2005. Quorum Sensing : Communication in Bacteria. *Annual Reviews in Cell Development Biology* **21**, 319–346.

Weitz D, Harder D, Casagrande F *et al.*, 2007. Functional and Structural Characterization of a Prokaryotic Peptide Transporter with Features Similar to Mammalian PEPT1. *Journal of Biological Chemistry* **282**, 2832–2839.

Xia Y, 2004. Proteases in pathogenesis and plant defence. *Cellular Microbiology* **6**, 905–913.

Xu R-Q, Blanvillain S, Feng J-X *et al.*, 2008. AvrAC(Xcc8004), a type III effector with a leucine-rich repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. *Journal of bacteriology* **190**, 343–355.

Yamaguchi K, Yamada K, Kawasaki T, 2013. Receptor-like cytoplasmic kinases are pivotal components in pattern recognition receptor-mediated signaling in plant immunity. *Plant Signaling & Behavior* **8**, e25662.

Zeier J, 2013. New insights into the regulation of plant immunity by amino acid metabolic pathways. *Plant, Cell and Environment* **36**, 2085–2103.

Zhang J, Chen L-L, Xing F *et al.*, 2016. Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. *Proceedings of the National Academy of Sciences* **113**, E5163–E5171.

Zhou J-M, Chai J, 2008. Plant pathogenic bacterial type III effectors subdue host responses. *Current opinion in microbiology* **11**, 179–185.

IV. Final remarks and future perspectives

Final remarks and future perspectives

The work presented in this thesis allowed the fulfillment of the proposed goals, described in the research strategy (Part I, Chapter 4), as further detailed below.

A polyphasic approach combining the determination of pathogenicity, virulence, population structure and phylogenetic diversity was used to identify and characterize a set of 33 *Xanthomonas campestris* isolates collected in Portugal. In addition to *X. campestris* pv. *campestris*, detected in the country in the 1960's, pathogenicity tests on a wide host range demonstrated the presence of pathovars *X. campestris* pv. *raphani* and *X. campestris* pv. *incanae* for the first time, in Portugal. Among *X. campestris* pv. *campestris* isolates, presence of races 4, 6 and 7 was recorded, and two novel races of this pathovar, race 10 and race 11, were also described. Phylogenetic relationships using *gyrB* gene confirmed phenotypic identification at species level and grouped *X. campestris* isolates in clades mostly congruent with pathovar allocation. Screening of eight essential virulence genes established five different partial virulence profiles, and the full set of genes was present in most isolates. When phylogenetic allocation and virulence profiles were compared, no evident associations with race, host and county of origin were found, confirming the high level of diversity that characterizes *X. campestris* pathovars in Portugal, and the lack of congruence between genomic and phenotypic features.

Using a MLST-based approach, phylogenetic diversity and population structure of a collection of 75 *X. campestris* isolates from several hosts and geographic regions was assessed. Although this population displayed a major clonal structure, neighbor-net phylogenetic analysis highlighted the presence of recombinational events that may have driven the ecological specialization of *X. campestris* and distinct host ranges. A high level of genetic diversity within and among pathovars of *X. campestris* was also revealed, through the establishment of 46 Sequence Types (STs). This approach was able to provide a snapshot of the worldwide *X. campestris* population structure, correlating the existing pathovars with three distinct genetic lineages.

The phylogenetic relationships between the founder genotype and the remaining isolates that constitute *X. campestris* pv. *campestris* populations were further clarified by the use of goeBURST algorithm. This analysis identified an intermediate link between *X. campestris* pv. *campestris* and *X. campestris* pv. *raphani* and disclosed new insights on the mechanisms driving the differentiation of both pathovars. The worldwide distribution of allelic variants suggests that *X. campestris* evolution as a seed-borne pathogen was not shaped by natural barriers. However, as Portuguese isolates encompass 26 unique STs, and this country is an important centre of domestication of *B. oleracea* crops, a strong case is made for its role as a diversification reservoir, most likely through host-pathogen co-evolution.

Although several genomic approaches have been used to characterize the molecular mechanisms of host-pathogen interaction, little was known about *X. campestris* pv. *campestris* regulation of virulence during pathogenesis. The RNA-Seq approach used to study the *in vivo* transcriptome profiling of two strains with contrasting virulence (L-vir and H-vir), inoculated on two cultivars of *B. oleracea*, allowed establishing that this pathogen undergoes transcriptional reprogramming in a host-independent manner, suggesting that virulence is an inherent characteristic. A total of 154 differentially expressed genes (DEGs) were identified in the contrasting virulent strains. The most represented functional category of DEGs was, as expected, 'pathogenicity and adaptation'. However, almost all functional categories were represented, highlighting the complex adaptation of pathogen cells to the host. In addition to Type III effectors, genes encoding proteins involved in signal transduction systems, transport, detoxification mechanisms and other virulence-associated processes were found differentially expressed between both strains.

These findings suggest that *X. campestris* pv. *campestris* does not rely on a restricted effector arsenal, but virulence seems to be the result of a fine-tuning mechanism, involving proteins from distinct cellular processes. This is composed of multiple layers and differences between strains appear to be the result of a complex regulation network, involving both virulence-related and basic cellular mechanisms. Overall, low virulence appears to be the combined result of impaired sensory mechanisms, reduced detoxification of reactive oxygen species (ROS), decreased motility and higher production of pathogen-associated molecular patterns (PAMPs), associated to an overexpression of avirulence proteins and a repression of virulence proteins targeting the hosts' PAMP-triggered immune responses. Using a highly innovative approach, the analysis of differential *in planta* infections allowed the disclosure of novel virulence related genes in *Xanthomonas campestris* pv. *campestris* - *B. oleracea* pathosystem, that will be crucial to further detail the virulence regulation network and develop new tools for the control of black rot disease. In this context, the validation of RNA-Seq results in the original sequenced samples by qPCR has to be carried out for target DEGs, selected on the basis of expression level, fold change and gene coverage. Biological validation of results should also be performed in independent biological samples of new pathosystems including *Brassica oleracea* cv. 'Beira' as a new host to validate the gene expression profiles of L-vir and H-vir strains. For this purpose, total RNA extraction, quantification and quality control for these pathosystems has already been carried out.

The establishment of the proteins directly involved in virulence regulation will allow for a more targeted and effective identification of their cognate host receptors. In such manner, the genetic background of race-nonspecific genes, conferring quantitative resistance, combined with race-specific resistance genes, conferring qualitative resistance, could potentially provide sustainable genetic control of black rot disease.

Appendix

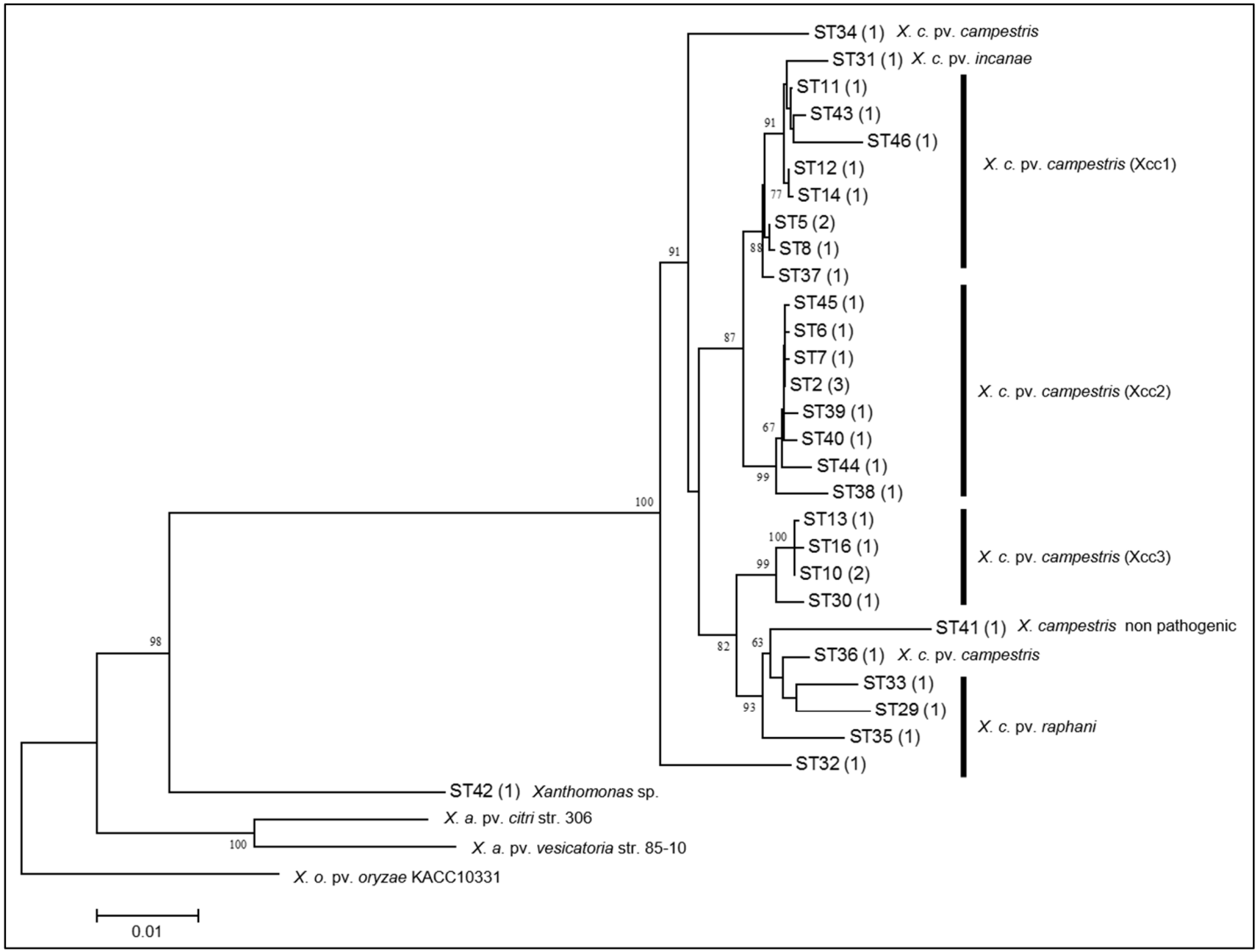


Figure 1 -Neighbor-Joining tree constructed with the concatenated sequences of *dnaK*, *fyuA*, *gyrB* and *rpoD* from the 33 Portuguese isolates. The number of isolates grouped in each sequence type (ST) is given in brackets. Bootstrap scores greater than 60 are displayed. For isolate designation and ST correspondence see Table 1 (Part II, Chapter 2). X. c. for *Xanthomonas campestris*; X. o. for *Xanthomonas oryzae*.

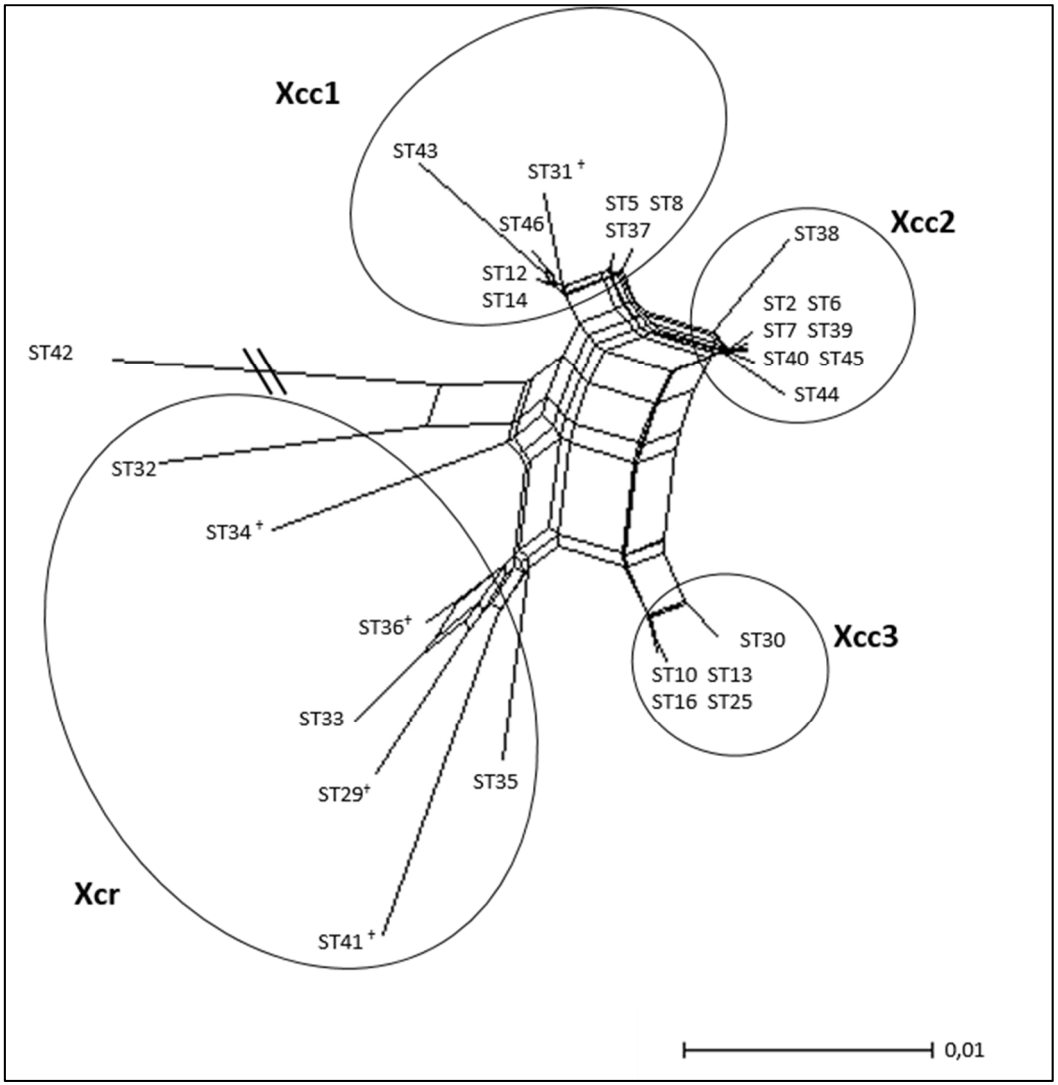


Figure 2 - Neighbor-net constructed with the concatenated sequences of *dnaK*, *fyuA*, *gyrB* and *rpoD* from the 33 Portuguese isolates. The designation at the branches indicates the ST number. Pathovars are indicated in the figure: Xcc – *X. campestris* pv. *campestris*; Xcr – *X. campestris* pv. *raphani*; Xci – *X. campestris* pv. *incanae*. † - isolates with identification different from the remaining ones in the cluster.

Table 1 – Virulence of *Xanthomonas campestris* pv. *campestris* strains against *B. oleracea* var. *sabauda* cv. 'Wirosa F1', measured as the mean ratio of infected leaf area (ILA) in old leaves (OL) and young leaves (YL) (mean ± standard deviation).

	ILA Old Leaves (OL)						ILA Young Leaves (YL)			
CPBF #	OL1	OL2	OL3	OL4	OL5	Total OL	YL1	YL2	YL3	Total YL
46	0.07	0.11	0.02	0.07	0.00	0.05 ± 0.04	0.17	0.21	0.09	0.16 ± 0.06
140	0.19	0.17	0.13	0.30	0.22	0.20 ± 0.06	0.32	1.00	1.00	0.77 ± 0.39
208	0.02	0.10	0.05	0.06	0.09	0.06 ± 0.03	1.00	1.00	1.00	1.00 ± 0.00
210	0.15	0.34	0.16	0.22	0.07	0.19 ± 0.10	0.26	0.15	1.00	0.47 ± 0.46
211	0.32	0.23	0.06	0.34	0.14	0.22 ± 0.12	0.19	0.26	0.24	0.23 ± 0.04
212	0.07	0.04	0.06	0.08	0.13	0.07 ± 0.03	0.17	0.21	0.06	0.15 ± 0.08
213	0.09	0.01	0.05	0.02	0.00	0.03 ± 0.03	0.01	0.03	0.04	0.02 ± 0.02
216	0.43	0.16	0.09	0.18	0.06	0.18 ± 0.15	0.40	1.00	1.00	0.80 ± 0.34
278	0.07	0.05	0.08	0.09	0.02	0.06 ± 0.03	1.00	1.00	1.00	1.00 ± 0.00
279	0.10	0.42	0.09	0.34	0.02	0.19 ± 0.18	0.13	0.20	1.00	0.44 ± 0.48
329	0.19	0.24	0.16	0.03	0.00	0.13 ± 0.10	0.08	0.22	0.07	0.12 ± 0.08
330	0.06	0.12	0.23	0.05	0.08	0.11 ± 0.07	0.13	0.03	1.00	0.38 ± 0.54
331	0.13	0.22	0.04	0.03	0.08	0.10 ± 0.08	1.00	1.00	1.00	1.00 ± 0.00
332	0.23	0.17	0.09	0.04	0.00	0.11 ± 0.10	0.24	0.09	0.23	0.19 ± 0.08
489	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
589	0.04	0.05	0.05	0.07	0.06	0.05 ± 0.01	0.35	1.00	1.00	0.78 ± 0.38
604	0.16	0.00	0.05	0.01	0.02	0.05 ± 0.07	1.00	1.00	1.00	1.00 ± 0.00
667	0.04	0.13	0.07	0.16	0.02	0.08 ± 0.06	0.18	0.29	0.47	0.31 ± 0.15
668	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
824	0.16	0.23	0.02	0.05	0.09	0.11 ± 0.09	0.28	0.35	0.19	0.27 ± 0.08
1126	0.05	0.19	0.35	0.19	0.06	0.17 ± 0.12	0.15	1.00	1.00	0.72 ± 0.49
1135	0.03	0.19	0.18	0.17	0.08	0.13 ± 0.07	0.23	0.23	0.42	0.29 ± 0.11
1136	0.05	0.12	0.09	0.09	0.06	0.08 ± 0.03	0.14	0.27	1.00	0.47 ± 0.46
1175	0.02	0.05	0.06	0.07	0.09	0.06 ± 0.03	0.07	0.07	0.17	0.10 ± 0.06
1176	0.12	0.11	0.24	0.06	0.02	0.11 ± 0.08	0.10	0.14	0.09	0.11 ± 0.03
1198	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00

Table 2 – Virulence of *Xanthomonas campestris* pv. *campestris* strains against *B. oleracea* var. *sabellica* cv. 'Bonanza F1', measured as the mean ratio of infected leaf area (ILA) in old leaves (OL) and young leaves (YL) (mean ± standard deviation).

	ILA Old Leaves (OL)						ILA Young Leaves (YL)			
CPBF #	OL1	OL2	OL3	OL4	OL5	Total OL	YL1	YL2	YL3	Total YL
46	0.04	0.12	0.07	0.05	0.00	0.06 ± 0.04	0.04	0.02	0.06	0.04 ± 0.02
140	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
208	0.01	0.26	0.10	0.11	0.00	0.10 ± 0.01	0.07	0.11	0.30	0.16 ± 0.13
210	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
211	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
212	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
213	0.05	0.04	0.11	0.05	0.10	0.07 ± 0.03	0.05	0.16	1.00	0.40 ± 0.52
216	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
278	0.05	0.18	0.25	0.16	0.01	0.13 ± 0.10	0.21	1.00	1.00	0.74 ± 0.45
279	0.09	0.15	0.03	0.06	0.00	0.06 ± 0.06	0.53	0.17	1.00	0.56 ± 0.42
329	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
330	0.10	0.03	0.01	0.13	0.00	0.05 ± 0.06	0.21	1.00	1.00	0.74 ± 0.45
331	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
332	0.06	0.03	0.07	0.24	0.00	0.08 ± 0.09	0.21	1.00	1.00	0.74 ± 0.45
489	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
589	0.02	0.23	0.17	0.17	0.21	0.16 ± 0.08	0.36	0.29	1.00	0.55 ± 0.39
604	0.23	0.06	0.10	0.20	0.11	0.14 ± 0.07	1.00	1.00	1.00	1.00 ± 0.00
667	0.14	0.16	0.17	0.09	0.03	0.12 ± 0.06	0.09	0.03	1.00	0.37 ± 0.54
668	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
824	0.04	0.03	0.09	0.03	0.09	0.06 ± 0.03	0.22	0.08	0.21	0.17 ± 0.08
1126	0.13	0.03	0.03	0.00	0.00	0.04 ± 0.05	0.38	0.24	0.10	0.24 ± 0.14
1135	0.03	0.13	0.04	0.01	0.00	0.04 ± 0.05	0.17	0.31	0.11	0.19 ± 0.10
1136	0.08	0.03	0.11	0.18	0.11	0.10 ± 0.05	0.06	0.06	0.04	0.06 ± 0.01
1175	0.31	0.04	0.07	0.06	0.20	0.14 ± 0.11	0.10	0.08	0.05	0.08 ± 0.02
1176	0.07	0.02	0.10	0.02	0.13	0.07 ± 0.05	0.11	0.10	0.15	0.12 ± 0.03
1198	0.01	0.05	0.02	0.07	0.00	0.03 ± 0.03	0.04	0.05	0.02	0.04 ± 0.02

Table 3 – Virulence of *Xanthomonas campestris* pv. *campestris* strains against *B. oleracea* var. *costata* cv. 'Beira F1', measured as the mean ratio of infected leaf area (ILA) in old leaves (OL) and young leaves (YL) (mean ± standard deviation).

	ILA Old Leaves (OL)						ILA Young Leaves (YL)			
CPBF #	OL1	OL2	OL3	OL4	OL5	Total OL	YL1	YL2	YL3	Total YL
46	0.01	0.07	0.06	0.02	0.05	0.04 ± 0.02	0.07	0.09	1.00	0.39 ± 0.53
140	0.14	0.19	0.01	0.06	0.00	0.08 ± 0.08	1.00	1.00	1.00	1.00 ± 0.00
208	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
210	0.26	0.03	0.16	0.00	0.00	0.09 ± 0.12	0.42	1.00	1.00	0.81 ± 0.33
211	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
212	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
213	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.02	0.02	0.10	0.05 ± 0.05
216	0.03	0.03	0.00	0.00	0.00	0.01 ± 0.02	0.22	0.13	1.00	0.45 ± 0.48
278	0.08	0.03	0.05	0.37	0.35	0.18 ± 0.17	1.00	1.00	1.00	1.00 ± 0.00
279	0.09	0.09	0.05	0.03	0.00	0.05 ± 0.04	0.42	1.00	1.00	0.81 ± 0.33
329	0.05	0.18	0.27	0.26	0.17	0.19 ± 0.09	0.42	1.00	1.00	0.81 ± 0.34
330	0.06	0.02	0.15	0.13	0.00	0.07 ± 0.07	0.52	0.34	0.27	0.38 ± 0.13
331	0.03	0.04	0.04	0.03	0.09	0.05 ± 0.03	1.00	1.00	1.00	1.00 ± 0.00
332	0.13	0.08	0.18	0.47	0.25	0.22 ± 0.15	0.36	0.35	1.00	0.57 ± 0.37
489	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
589	0.23	0.09	0.18	0.05	0.22	0.15 ± 0.08	0.28	0.14	0.30	0.24 ± 0.09
604	0.02	0.02	0.05	0.04	0.46	0.12 ± 0.19	0.26	0.25	0.18	0.23 ± 0.04
667	0.11	0.29	0.33	0.17	0.17	0.21 ± 0.10	0.17	0.25	1.00	0.47 ± 0.46
668	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00
824	0.34	0.13	0.15	0.24	0.07	0.19 ± 0.11	0.31	0.03	0.10	0.15 ± 0.15
1126	0.02	0.08	0.18	0.43	0.42	0.23 ± 0.19	0.32	0.27	0.54	0.38 ± 0.14
1135	0.10	0.10	0.05	0.35	0.37	0.19 ± 0.15	0.41	0.30	0.26	0.32 ± 0.07
1136	0.10	0.09	0.19	0.42	0.10	0.18 ± 0.14	0.14	0.52	0.45	0.37 ± 0.20
1175	0.19	0.14	0.11	0.29	0.55	0.26 ± 0.18	0.20	0.25	1.00	0.48 ± 0.45
1176	0.19	0.19	0.13	0.10	0.33	0.19 ± 0.09	0.40	0.50	0.09	0.33 ± 0.22
1198	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	0.00	0.00	0.00	0.00 ± 0.00